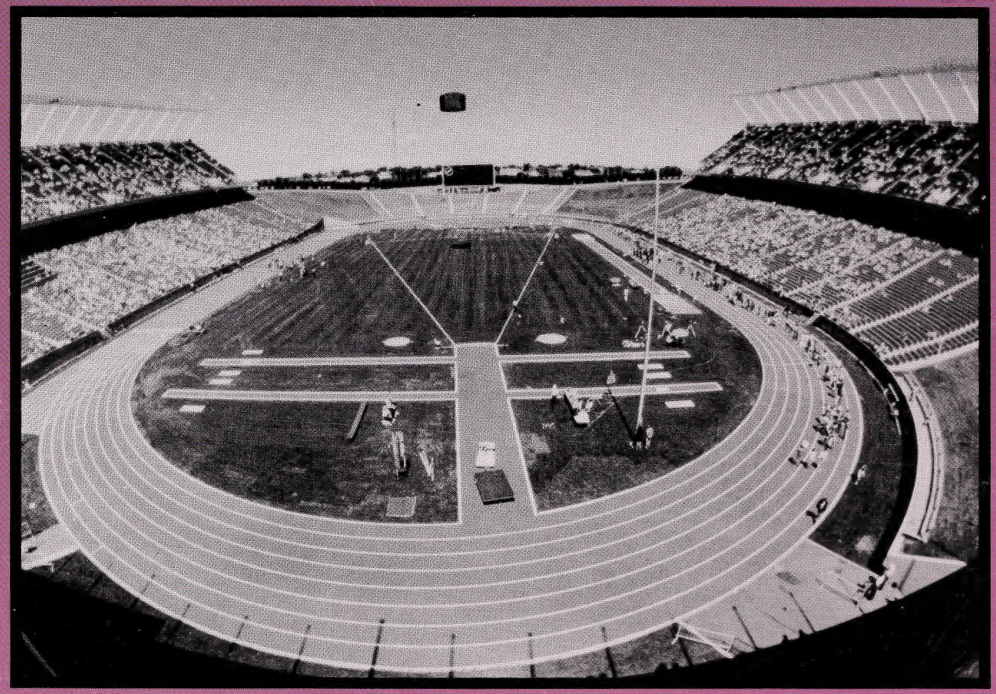
MATHEMATICS 10

University of Alberta Library
0 1620 3452378 5

UNIT 8



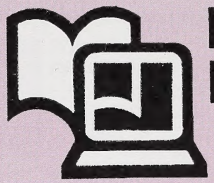
STATISTICS

Distance Learning

Alberta EDUCATION

UNIVERSITY LIBRARY

Welcome



Distance Learning

You have chosen an alternate form of learning that allows you to work at your own pace. You will be responsible for your own schedule, for disciplining yourself to study the units thoroughly, and for completing your units regularly. We wish you much success and enjoyment in your studies.

ALL RIGHTS RESERVED

Mathematics 10 Student Module Unit 8 Statistics Alberta Distance Learning Centre ISBN No. 0-7741-0755-3

Copyright © 1992, the Crown in Right of Alberta, as represented by the Minister of Education, Alberta Education, 11160 Jasper Avenue, Edmonton, Alberta, T5K 0L2. All rights reserved. Additional copies may be obtained from the Learning Resources Distributing Centre.

No part of this courseware may be reproduced in any form, including photocopying (unless otherwise indicated), without the written permission of Alberta Education.

Every effort has been made both to provide proper acknowledgement of the original source and to comply with copyright law. If cases are identified where this has not been done, please notify Alberta Education so appropriate corrective action can be taken.

Digitized by the Internet Archive
in 2016 with funding from
University of Alberta Libraries

https://archive.org/details/mathematics1008albe

General Information

This information explains the basic layout of each booklet.

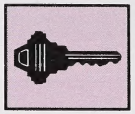
- **What You Already Know** and **Review** are to help you look back at what you have previously studied. The questions are to jog your memory and to prepare you for the learning that is going to happen in this unit.
- As you begin each **Topic**, spend a little time looking over the components. Doing this will give you a preview of what will be covered in the topic and will set your mind in the direction of learning.
- **Exploring the Topic** includes the objectives, concept development, and activities for each objective. Use your own papers to arrive at the answers in the activities.
- **Extra Help** reviews the topic. If you had any difficulty with **Exploring the Topic**, you may find this part helpful.
- **Extensions** gives you the opportunity to take the topic one step further.
- To summarize what you have learned, and to find instructions on doing the unit assignment, turn to the **Unit Summary** at the end of the unit.
- The **Appendices** include the solutions to **Activities (Appendix A)** and any other charts, tables, etc. which may be referred to in the topics **(Appendix B, etc.)**.

Visual Cues

Visual cues are pictures that are used to identify important areas of the material. They are found throughout the booklet.

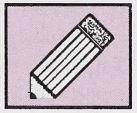
An explanation of what they mean is written beside each visual cue.

What You Already Know
- reviewing what you already know



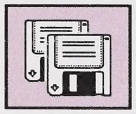
Key Idea
- flagging important ideas

Audiotape
- learning by listening to an audiotape



Review
- studying previous concepts

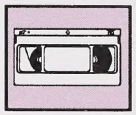
Another View
- exploring different perspectives



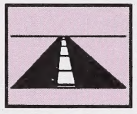
Computer Software
- learning by using computer software

Introduction
- introducing the unit

Solutions
- correcting the activities



Videotape
- learning by viewing a videotape



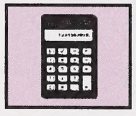
What Lies Ahead
- previewing the unit

Extra Help
- providing additional study

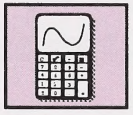
Print Pathway
- choosing a print alternative

Exploring the Topic
- actively learning new concepts

Extensions
- going on with the topic



Calculator
- using your calculator



Graphing Calculator
- using your graphing calculator

What You Have Learned
- summarizing what you have learned

Mathematics 10

Course Overview

Mathematics 10 contains 8 units. Beside each unit is a percentage that indicates what the unit is worth in relation to the rest of the course. The units and their percentages are listed below. You will be studying the unit that is shaded.

Unit 1 Number Systems	10%
Unit 2 Operations on Polynomials	14%
Unit 3 Equations and Inequalities	10%
Unit 4 Factoring Polynomials	13%
Unit 5 Coordinate Geometry	19%
Unit 6 Systems of Equations	10%
Unit 7 Trigonometry	11%
Unit 8 Statistics	13%
	100%

Unit Assessment

After completing the unit, you will be given a mark based totally on a unit assignment. This assignment will be found in the Assignment Booklet.

Unit Assignment - 100%

If you are working on a CML terminal, your teacher will determine what this assessment will be. It may be

Unit Assignment - 50%
Supervised Unit Test - 50%

Introduction to Statistics

This unit covers topics dealing with statistics. Each topic contains explanations, examples, and activities to assist you in understanding statistics. If you find you are having difficulty with the explanations and the way the material is presented, there is a section called **Extra Help**. If you would like to extend your knowledge of the topic, there is a section called **Extensions**.

You can evaluate your understanding of each topic by working through the activities. Answers are found in the solutions in **Appendix A**. In several cases there is more than one way to do the question.

Unit 8 Statistics

Contents at a Glance

Value		
	Statistics	3
	What You Already Know	5
	Review	6
15%	**Topic 1**: Collection of Data	8
	• Introduction • Extra Help • What Lies Ahead • Extensions • Exploring Topic 1	
30%	**Topic 2**: Organization and Presentation of Data	25
	• Introduction • Extra Help • What Lies Ahead • Extensions • Exploring Topic 2	
30%	**Topic 3**: Analysis of Data: Mean, Median, Mode	82
	• Introduction • Extra Help • What Lies Ahead • Extensions • Exploring Topic 3	
25%	**Topic 4**: Drawing Inferences	100
	• Introduction • Extra Help • What Lies Ahead • Extensions • Exploring Topic 4	
	Unit Summary	118
	• What You Have Learned • Unit Assignment	
	Appendices	119
	• Appendix A • Appendix B	

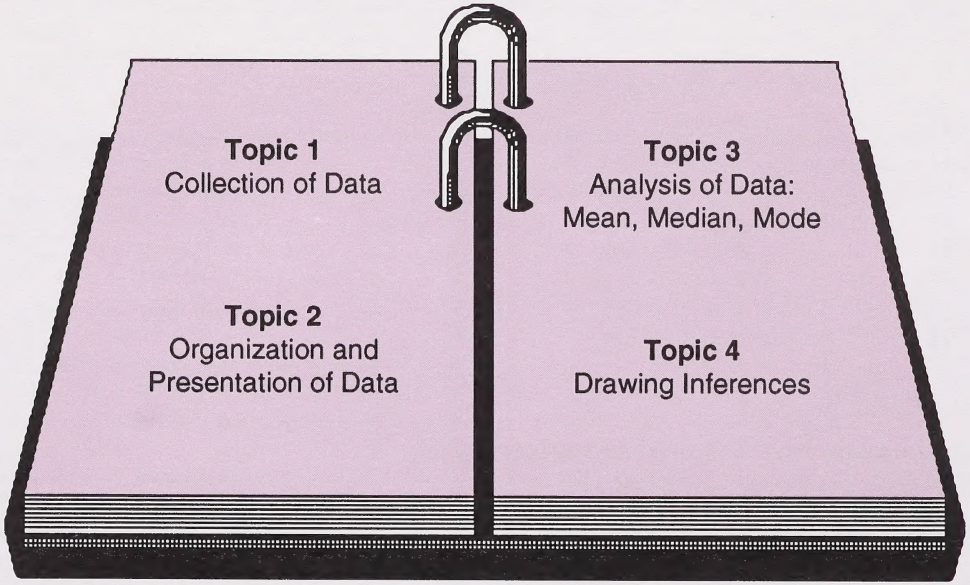
Statistics

Statistics influence your daily life. Through statistics, you can elect governments; measure intelligence, interests, aptitudes; determine which television shows will survive; compare performances of athletes; decide which business strategy to follow; decide whether it will rain tomorrow; choose the top twenty songs; and keep informed about what is going on in the world.

Top 20 Hits

1. Lovebird in the Jungle
2. O Baby
3. Down the Road
4. Ragtime Pinks
5. Crying In The Sunshine
6. Hit the Beach
7. Purple Scarves
8. Beach Ball Blues
9. Dirty Dan
10. Yellow Corvette
11. Counting Raindrops
12. Umbrella Bella
13. Mackintosh Melody
14. Sandra's Sweetheart
15. Tony's Lovebug
16. Laserwriter Blues
17. Paper Heart
18. Keep Off The Grass
19. Glass Cabin
20. Carpet Burner

Unit 8 Statistics

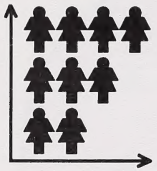


What You Already Know

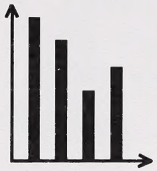
Refresh your memory!

You should be able to do the following:

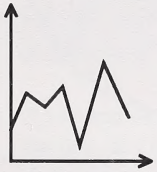
- Understand the purpose of statistics, recognize the use and misuse of statistics, and detect biases in surveys.
- Represent data in the form of pictographs, bar graphs, line graphs, and circle graphs.



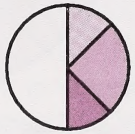
pictograph



bar graph



line graph



circle graph

- Analyse and interpret arguments or conclusions based on statistical information. For example, if only 14% of those surveyed approved of the current government, it may not get re-elected.
- Understand and use the term **average** when dealing with data from meaningful situations.

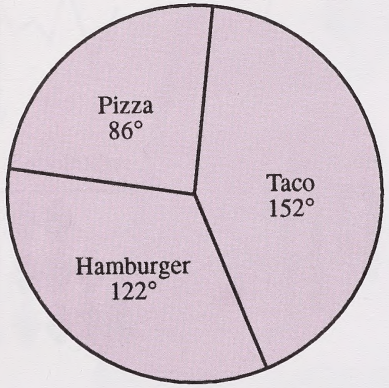
  For example, if you are given 2, 4, 9, 17, and 23, the average is calculated as follows:

$$\text{Average} = \frac{2+4+9+17+23}{5}$$
$$= \frac{55}{5}$$
$$= 11$$

- Conduct a survey or a poll using correct sampling techniques, and report results using an appropriate table, chart, or graph.

For example, fifty students at Griffiths High School were asked which fast food they preferred. The responses were recorded as follows:

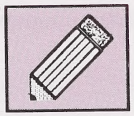
Hamburger	Pizza	Taco
17	12	21



- Understand and use the term **probability**. Probability refers to the likelihood of an event happening. For example, the Edmonton Eskimos will probably win the Grey Cup next year.

- Express the probability of the occurrence of an event from a practical situation or a simple experiment or simulation. For example, the probability of getting heads when you flip a coin is $\frac{1}{2}$.

Now that you have looked at material you studied previously, go to the **Review** to confirm your understanding of this material.



Review

Now try the following questions.

1. Put the following information into a pictograph.

Number of Blocks Travelled to School

Person	Blocks
Jane	5
Dick	10
Audrey	7
Jimmy	2

2. In a small lake, 50 fish are caught, tagged, and then released back into the lake. During the following year, a total of 200 fish are caught. Of these 200 fish, 25 were fish that had been tagged. The suppliers to the lake then predict that there are still another 200 fish in the lake. Do you agree with this prediction? Support your answer.

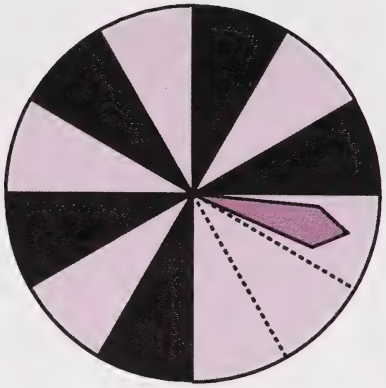
3. What is the average of 7, 10, 13, 16, and 19?

4. What type of graph should you use to report the change in temperature over a twenty-four-hour period?

   A. bar graph
   B. circle graph
   C. line graph
   D. pictograph

5. The Blue Rockets won fourteen of their last sixteen games. What is the probability that they will win their next game?

6. A spinner has a twelve pie-shaped pieces that are equal in size. If seven pieces are red and five pieces are black, what is the probability that the spinner will stop on a red piece?



Now go to the **Review** solutions in **Appendix A**.

If you had difficulties with these questions, you may need to review the unit on **Data Management** in **Mathematics 9** for a more extensive review before you start this unit.

Topic 1 Collection of Data

Introduction

If members of a student council want to hold a dance, how can they be sure that it will be the type of dance that most students want? How would collect the information needed? Some different possibilities are listed as follows:

- They can ask every student in school.
- They can have students return surveys from the school newspaper.
- They can look over previous dance profits to see which were successful.
- They can survey some of the students in the cafeteria.
- They can get a list of all the students, choose some randomly, and phone them for their opinions.

What Lies Ahead

Throughout the topic you will learn to

1. analyse problems and generate and collect data appropriately
2. identify samples along with the populations from which they are taken
3. evaluate samples for bias, appropriateness of sample type, and randomness

Now that you know what to expect, turn the page to begin your study of collection of data.

Exploring Topic 1

Activity 1

Analyse problems and generate and collect data appropriately.

You are constantly exposed to information. It may include data which you have actively collected. This information could be fact or it could be opinion.

You will find it important at times to test your belief or opinion about something.

For example, you may have witnessed two serious traffic accidents at a particular uncontrolled intersection. You may believe this is a dangerous intersection for you to use.

It is possible that these accidents were two of a large number that occurred there during the past year. It is also possible that you observed the only two accidents at that intersection during the past year.

In order to verify or reject your belief, you will need to collect more information (called data) on traffic accidents at this intersection, organize and analyse the data, and interpret the data in relation to your original belief.

Different methods of collecting data can be used. In the previous example, you may wish to collect data from past records regarding accidents at that particular intersection.

Methods of collecting data about other situations could include the following:

- personal interviews, questionnaires, or polls
- experiments designed to collect data

The first step involved in statistics is illustrated by the following.

For example, a manufacturer of compact laser discs (CDs) wishes to check their quality before they are sent to the retail distributor.

The manufacturer has the option of testing each CD. Taking a census like this would not be in the best interest of the manufacturer. This would be a time-consuming and expensive option.

In the interests of saving time and money, the manufacturer may select a **sample** from the CDs.

In order for the manufacturer to receive correct results from the sample, the following must occur.

- The sample must be **representative** of the population.
- The sample must be selected at **random**.

The following are important statistical terms. Examine these terms carefully.

- Data: factual information which is used for calculations, discussions, or reasoning
- Census: use of all items of the population to gather the data
- Population: the whole number of people or items for which data is being collected
- Sample: a part of the population that is used to study the characteristics of the population
- Random sample: a sample selected such that each member of the population has an equal chance of being selected for the sample
- Biased: a different outcome of the sample compared to the outcome of the population

Do the following questions.

1. The following are methods of data collection.

   - telephone poll
   - questionnaire
   - experiment
   - personal interview
   - research records

   Which method of data collection should be used for the following? There may be more than one answer.

   a. What improvements should be made to the recreation centre?

   b. What is the average temperature for each province?

   c. Which is the most popular political party?

   d. How many people enjoyed a particular movie?

   e. What do students eat for breakfast?

   f. Which weeds does a new weed killer destroy?

2. For each of the following situations, should a census or a sample be used?

   a. Check the quality of a shipment of lettuce.

   b. Check the quality of parachutes that are being manufactured.

c. You wish to know how the public will vote for the candidates in a local election.

d. A forest ranger wishes to check to see if a stand of timber is ready to be harvested.

For solutions to **Activity 1**, turn to **Appendix A, Topic 1**.

Activity 2

Identify samples along with the populations from which they are taken.

Collecting data from a sample instead of a population is a time- and money-saving measure. However, this effort will be wasted if the results from the sample are not similar to what would be obtained from the population. For this reason, great care must be taken in the selection of the sample. To make sure that the sample gives dependable results, the sample must be large enough and the sample must be representative of the population.

Sample Size

The size of the sample you select will depend on the following two things:

- the subject of the sample
- the size of the population

The more variations that are present in a population, the larger the sample you will require. This can be seen if you are surveying a community. If the community is made up of a large number of different ethnic groups, you will have to increase the size of your sample. This will ensure that you are getting the opinions of all of the ethnic groups instead of a community that has only one ethnic group.

The size of the population is very important in determining the size of the sample. There is no set rule that can be used to decide on the size of the sample.

The general principle for deciding on the size of the sample is as follows: the larger the population, the larger the sample.

The following chart gives an indication of how the size of a sample compares to the size of a population.

The Size of Samples Used by Arnie's Sound

Size of Population	10	100	1000	10 000	100 000	1 000 000	10 000 000
Size of Sample	10	100	750	5000	10 000	50 000	200 000
Percentage of Population	100%	100%	75%	50%	10%	5%	2%

This chart illustrates that when the population is small, you should take a census instead of a sample. As the population get larger, the sample becomes a smaller portion of the population.

This chart does not give the sample sizes that should be used for a population. It is only used here as an indicator. As noted before, the characteristics of the population must also be taken into consideration when deciding on the size of the sample.

For example, a sample for a population of 500 000 should be from 5% to 10% of the population since 500 000 is between 100 000 and 1 000 000 on the chart.

Since 5% of 500 000 is 25 000 and 10% of 500 000 is 50 000, the sample should be at least 25 000 and at most 50 000.

Sample Representation

The characteristics of the sample should be in proportion to the characteristics of the population.

What does this mean? To answer this, return to the case where a manufacturer of compact laser discs is checking the quality of the CDs.

The following chart shows how many CDs were made and how many are used in the sample for each artist. Since the total population is 2 500 000, the chart requires a sample of 2% to 5%. A sample of 3% is chosen.

Actual Sample Sizes Used For Testing by Arnie's Sound

Artist	Size of Population	Size of Sample	Percentage of Population
Belinda Aisle	300 000	9 000	3%
Brian Odums	150 000	4 500	3%
Dadonna	400 000	12 000	3%
Flatwood Joe	250 000	7 500	3%
Men Out of Work	125 000	3 750	3%
Neil Zirconia	175 000	5 250	3%
U3$\frac{1}{2}$	750 000	22 500	3%
Von Jodi	350 000	10 500	3%
TOTAL	2 500 000	75 000	3%

In this case the manufacturer is taking the same proportion or percentage of each artist's CDs for the sample. For each artist, the manufacturer is taking 3% of the CDs and studying them for defects.

Do any three of the following questions.

1. Using the chart, decide what size of sample to use for each of the following population sizes?

Size of Population	10	100	1000	10 000	100 000	1 000 000	10 000 000
Size of Sample	10	50	250	1500	10 000	50 000	200 000
Percentage of Population	100%	50%	25%	15%	10%	5%	2%

   a. 20 000 b. 250

   c. 900 000 d. 5 000 000

2. Given the following sizes of a sample, what size of a population would you expect? Refer to the chart in question 1.

   a. 65 b. 2000

   c. 150 000 d. 35 000

3. Which of the following are reasonable sizes for the sample? Explain why any particular sample is not reasonable. Refer to the chart in question 1.

   a. A sample of 20 from a population of 500

   b. A sample of 50 000 from a population of 1 500 000

   c. A sample of 5000 from a population of 50 000

   d. A sample of 200 from a population of 250 000

4. Complete the following chart. Assume that the proportions of all the populations are the same as that for Trait 1.

	Population Size	Sample Size	Proportion of Population
Trait 1	100 000	5000	
Trait 2		7000	
Trait 3	750 000		
Trait 4		5500	
Total	1 100 000		

For solutions to **Activity 2**, turn to **Appendix A**, **Topic 1**.

Activity 3

Evaluate samples for bias, appropriateness of sample type, and randomness.

A sample is biased if all members of the population do not have an equal chance of being in the sample. For example, if you want an opinion from high school students but only interview grade tens, you will have a biased sample.

To avoid a biased sample, take a sample that is large enough and make some provision for all members of the population being represented.

This section explores the advantages and disadvantages of different sampling methods.

- You may choose a sample because it is readily available. For example, if you wanted to sample the opinion of students about the last dance, you could choose the first five students you see. This is called **convenience sampling**.

- When people participate in a survey by voluntarily returning a form printed in a newspaper, they make up a **self-selected sample**. However, people who care enough to respond may not be representative of the whole population.

- In another form of sampling, an expert selects a sample that is considered representative. For example, a produce buyer might select and taste several apples from a shipment in order to determine the quality of the shipment of apples as a whole. If the expert is really good at selecting the sample, a **judgement sample** may give a better estimate than most random samples would.

- A common type of sampling is **stratified sampling**. In this method, the population is divided into separate strata, or specified smaller groups. Then, a random sample in each group is taken. For example, to obtain a sample of forty students, you could divide the students into the two strata of boys and girls and take a random sample of twenty from each. This is done because you may want to know how the views of boys and girls compare. Thus, you must be certain to have enough boys and girls in the sample so you can define them as the strata.

- If you were to select a sample of students from your class by choosing every fifth student who walks into the classroom, you would be using **systematic sampling**.

Different types of sampling include the following:

- convenience
- self-selected
- judgement
- stratified
- systematic
- clustered
- destructive

- If you were to take a sample from a particular segment of the population, the sample is said to be **clustered**. For example, a manufacturer of tools for mechanics wants to know which magazines mechanics read, so the manufacturer will know where to place advertisements. In this case the manufacturer would only survey mechanics instead of the entire population.

- Sometimes a sample is damaged so much that it cannot be put back into the population. This type of sampling is called **destructive sampling**.

Look at the following example.

Example 1

The magazine *Women in Business* recently conducted a survey of 1100 readers and made the following comment: More than half of the secretaries in Canada would rather work for a man than a women, even though the man was more likely to ask them to make coffee.

From the 1100 readers surveyed, 68% prefer to work for a man, 8% prefer to work for a woman, and 24% did not have a preference.

- What type of sampling was done to get these results?

  Solution:

  The article does not give enough information to determine the sampling method, but it does **imply** that the sample was self-selected. The sample could also be considered to be clustered.

- To what population do these results apply?

  Solution:

  These results apply to all Canadian secretaries.

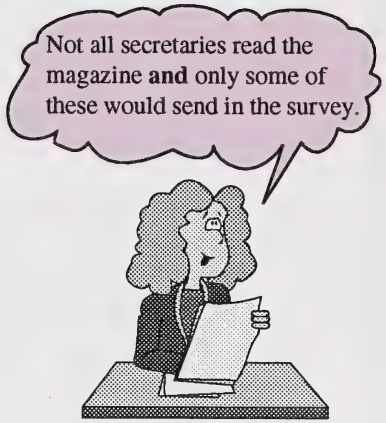
- State how this sampling might be biased.

  Solution:

  The survey is not likely to include male secretaries because the magazine, *Women in Business*, would probably appeal to women only.

In most samplings, you must make sure that all members of the population have an equal chance of being selected; this results in an **unbiased sample**.

One way to make sure that you pick a random sample is to use a random number table, like the one that follows. A more complete random number table is found in **Appendix B**.



Random Number Table

97954	71016	47792	57802	68548
08278	87607	89996	76407	99493
49006	90538	12392	60782	08097
84129	65136	99203	14518	40414
78118	44330	77823	69900	63633
89082	30031	34107	39437	51844
90941	72864	69974	06088	70920
51212	79280	58654	61752	63349
85756	53922	39478	18502	00688
85937	46714	37911	04118	34653

Suppose you want to select a sample of six names out of the following population to respond to a questionnaire. You may use a random number table to obtain the samples. Each item in the population can be numbered as shown in the following table. Other numbering systems can also be chosen and used if desired.

01	Bob	10	Greg	19	Murriel	28	Sarah
02	Bart	11	Harvey	20	Nora	29	Tabitha
03	Brad	12	Karen	21	Nancy	30	Toni
04	Betty	13	Kathy	22	Paul	31	Teresa
05	Donna	14	Kevin	23	Peter	32	Terrance
06	Dawn	15	Kristin	24	Pablo	33	Trevor
07	Frank	16	Keith	25	René	34	Victoria
08	Filbert	17	Linda	26	Robin	35	Wendall
09	Garth	18	Lois	27	Susan	36	Wendy

Random Number Table

97954	71016	47792	57802	68548
08278	**87607**	**89996**	**76407**	**99493**
49006	**90538**	**12392**	**60782**	**08097**
84129	65136	99203	14518	40414
78118	44330	77823	69900	63633
89082	30031	34107	39437	51844
90941	72864	69974	06088	70920
51212	79280	58654	61752	63349
85756	53922	39478	18502	00688
85937	46714	37911	04118	34653

Starting at any point on the random number table, pick two digits at a time. You may want to start at the second row and the second column. The first two digits are 87, 60 would be next, 78 after that, and so on until you have six numbers that are listed beside the names in the table. You cannot use numbers greater than 36 because there are only 36 names in the population. Here are the numbers that are selected if this starting point is used. Some other starting point would result in some different choices. Some choices may be the same no matter what the starting point is.

87	60	78	99	96	76	40	79	94	93	49
80	69	**05**	38	**12**	39	**26**	**07**	82	**08**	**09**

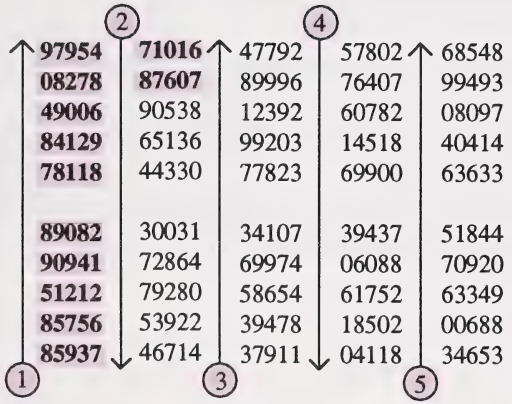
The sample consists of 05, 12, 26, 07, 08, and 09. The others are discarded because they are greater than 36. The following individuals would serve as the sample for your questionnaire: Donna, Karen, Robin, Frank, Filbert, and Garth.

See the **Extra Help** section for more on random number tables. This is not the only way to use a random number chart. You can use the numbers going up or down, right to left, or the first few digits from each group.

An important rule when using the table is to make sure that you use the same procedure to pick all of the numbers.

Suppose you were asked to choose ten people to serve on a discussion panel. Using the same table as before, your choices could be made as shown in Table 1. Whatever pattern is chosen, make sure to follow through with this same pattern until all ten people are chosen. Do not mix patterns as illustrated in Table 2.

Table 1

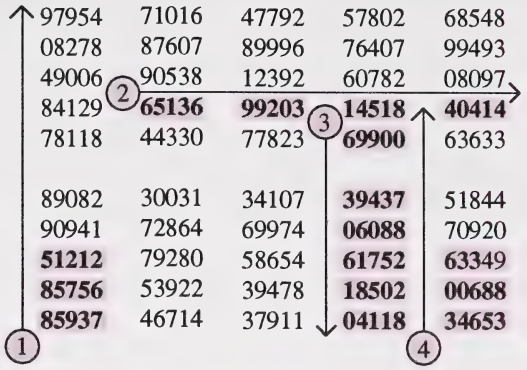


This is proper since the same pattern is followed until ten names are selected.

The numbers would be 85, 93, 78, 57, 56, 51, **21**, **29**, **09**, 41, 89, **08**, **27**, 81, **18**, 84, **12**, 94, 90, **06**, 08, 27, 89, 79, 54, 71, **01**, 68, 76, **07**.

The chosen people would be Nancy, Tabitha, Garth, Filbert, Susan, Lois, Karen, Dawn, Bob, and Frank.

Table 2



This is improper since the pattern changes to make a selection of ten names.

In this case the numbers would be 85, 93, 78, 57, 56, 51, **21**, **26**, 51, **36**, 99, **20**, **31**, 45, **18**, 40, 41, 46, 99, 00, 39, 43, 70, 60, 88, 61, 75, 21, 85, **02**, **04**, **11**, 83, 46, 53, 00, 68, 86, **33**.

Now try the following questions.

1. What type of sample is being collected in the following cases?

   a. A television reporter is interviewing people on a downtown street corner.

   b. A polling company selects every hundredth person from the telephone book and asks which radio station they are listening to.

   c. A tire company produces batches of 100 tires. Five tires from every batch are selected randomly and tested for quality.

2. Explain how each of the following cases is biased.

   a. The first thirty-five radios produced each day are tested for their quality.

   b. A local newspaper interviews people in front of the old town swimming pool about the construction of a new pool.

   c. A company selling a new hand soap tested it in three Edmonton malls to see how well it would sell across Canada.

3. In which of the following situations would destructive sampling be used?

   a. the life expectancy of light bulbs

   b. whether airplane engines can withstand certain stresses

   c. whether golf balls have the proper compression

   d. Whether a computer is properly set up

   e. Whether the correct percentage of vegetable seed will germinate

4. Discuss the advantages and disadvantages of using the following methods to gather a market analysis from consumers about certain products.

   a. door-to-door interview

   b. telephone poll

   c. mail-in questionnaire

5. Explain how you would get a representative sample from each of the following situations.

   a. a 15% sample from a high school with a population of 5000

   b. a 2% sample of a city with a population of 60 000

   c. a 1% sample of the motorcycle owners in Calgary

For solutions to **Activity 3**, turn to **Appendix A**, **Topic 1**.

If you require help, do the Extra Help section.

If you want more challenging explorations, do the Extensions section.

} You may decide to do both.

Extra Help

You may study this section by doing **Part A** or **Part B** or both. **Part A** consists of an audiotape while **Part B** is the print section. Whichever way you choose to study this section, complete the questions at the end of the print section.

Part A

Audio Activity

Insert the audiotape titled *Mathematics 10 – Random Number Tables* into your tape recorder and follow the instructions on the tape.

1

77160	13396	97957	38623	96977
66545	15729	27928	46327	38631
71748	57654	10000	90955	55843
25581	73139	60361	98089	24619
56352	06504	95647	74485	47450
51315	67337	66031	97841	87584
79526	05997	87262	66542	71412
67936	62421	21662	40143	15528
99895	48802	65542	88966	38133
71395	64846	46572	02491	94737

2

01 Arthur	12 Henry	23 Peter
02 Andrew	13 Henrietta	24 Pablo
03 Angie	14 Harvey	25 Robert
04 Barry	15 Kathleen	26 René
05 Brad	16 Kristin	27 Sarah
06 Betty	17 Laura	28 Tabitha
07 Carla	18 Lawrence	29 Teresa
08 Carol	19 Mark	30 Trevor
09 Carmen	20 Marilyn	31 Victoria
10 David	21 Nora	
11 Frank	22 Nolan	

3

77160	13396	97957	38623	96977
66545	15729	27928	46327	38631
71748	57654	10000	90955	55843
25581	73139	60361	98089	24619
56352	06504	95647	74485	47450
51315	67337	66031	97841	87584
79526	05997	87262	66542	71412
67936	62421	21662	40143	15528
99895	48802	65542	88966	38133
71395	64846	46572	02491	94737

4

77160	13396	97957	38623	96977
66545	15729	27928	46327	38631
71748	57654	10000	90955	55843
25581	73139	60361	98089	24619
56352	06504	95647	74485	47450
51315	67337	66031	97841	87584
79526	05997	87262	66542	71412
67936	62421	21662	40143	15528
99895	48802	65542	88966	38133
71395	64846	46572	02491	94737

5

91237	20271	07365	23090	53974
22499	85467	01343	01917	42677
38934	21232	88668	39984	48445
99045	57734	36677	76232	88758
22568	83371	27519	61790	22472
54397	13462	64815	72886	50159
56640	12431	92943	32903	83663
43814	75098	68176	80417	61984
03082	07036	73327	12751	84966
16645	20078	03882	21707	84231
28898	75365	10862	67819	96110
40211	41272	84515	53579	70146
93779	69626	93033	12884	61036
70865	33189	34901	26875	03086
47097	38770	33245	11617	73057

6

91237	20271	07365	23090	53974
22499	85467	01343	01917	42677
38934	21232	88668	39984	48445
99045	57734	36677	76232	88758
22568	83371	27519	61790	22472
54397	13462	64815	72886	50159
56640	12431	92943	32903	83663
43814	75098	68176	80417	61984
03082	07036	73327	12751	84966
16645	20078	03882	21707	84231
28898	75365	10862	67819	96110
40211	41272	84515	53579	70146
93779	69626	93033	12884	61036
70865	33189	34901	26875	03086
47097	38770	33245	11617	73057

Part B

More About Random Number Tables

In this section you are going to learn more about how to use random number tables.

Suppose Arnie wants to test a sample of 15 CDs out of a production of 100 CDs. He may use the random number table to choose his sample because the CDs are numbered 00 through 99.

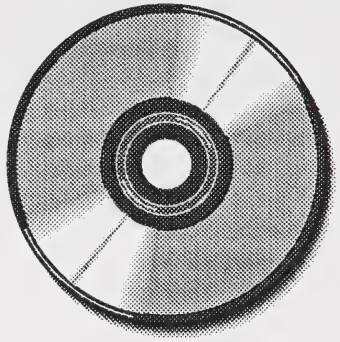
The following is part of a random number table. Arnie may begin at any number. Suppose he decides to start at 29269 and takes two digits at a time.

37876	**29269**	**19725**	89**11**2	5**13**15	**67**337	66**03**1	97841	87584
60441	25374	28623	68032	79526	05997	87262	66542	71412
04699	27298	28814	83771	67936	64241	21662	40143	15528
43248	64464	13875	62605	99895	48802	65542	88966	38133
19786	24621	29140	80520	71395	64846	46572	02491	94737

Every second pair of digits has been made bold so it is easier to read the pairs of digits.

His samples would consists of 29, 26, 91, 97, 25, 89, 11, 25, 13, 15, 67, 33, 76, 60, and 31.

He would test the CDs which are numbered according to the numbers selected.



The digits for the random number table can be selected in other ways. You might try using the reverse method of what was done here. Move from the right to the left and then up and to the right. As long as you use a consistent pattern when selecting the numbers, you should get randomly generated numbers.

37876	24269	19725	89112	51315	67337	66031	97841	87584
60441	25374	28623	68032	79526	05997	87262	66542	71412
04699	27298	28814	83771	67936	64241	21662	40143	15528
43248	64464	13875	62605	99895	48802	65542	88966	38133
19786	24621	29140	80520	71395	64846	46572	02491	94737

If the reverse method is used, Arnie's sample would consist of 73, 74, 91, 94, 20, 27, 56, 46, 48, 59, 31, 70, 25, 08, and 04, or else the samples would be 37, 47, 19, 49, 02, 72, 65, 64, 84, 95, 13, 07, 52, 80, and 40.

Notice that when the numbers 46 and 64 were repeated, they were left out. This is because a disc can be selected only once.

This clearly shows that there are many ways to select a sample from a random number table.

Suppose you had a random number table set up like as follows:

Notice the consistent pattern. Begin in the fourth column and go down. **Shade** the first three, then skip three. Keep doing this until the bottom of the fourth column is reached. Then go to the top of the fifth column and again shade three, then skip three. This sample would be made up of three-digit numbers.

35707	77160
86512	66545
72986	71748
40407	25581
97623	56352
99631	13396
85588	15729
95059	57654
21448	73139
03970	06504
50314	97957
09766	27928
07270	10000
04852	60361
95838	95647
74845	38623
52143	46327
18649	90955
94168	98089
14590	74485
85466	96977
71897	38631
32027	55843
72872	24619
28791	47450

The three-digit numbers are as follows: 018, 023, 854, 716, 753, 444, 696, 927, 964, 485, 925, 305, 206, 422, 588, 734, 157, 267, 318, 980, 460, 285, and 398.

Now try the questions that follow.

Do either question 1 or 2 and question 3.

1. Select any two rows of numbers from the random number table found in **Appendix B**. Put the digits into groups of four. What is the average of the twenty-five selected numbers?

2. Select any column from the random number table. Take the first three digits from each group of five. What is the average of the fifty selected numbers?

3. A population of 500 000 discs was made before quality testing was started by the manufacturer.

   a. If the manufacturer wants to test 0.1% of the discs, how many discs wil be tested?

   b. How many digits do you need to represent a disc if the discs are numbered 1 through 500 000?

   c. How many rows would you need on the random number table?

For solutions to **Extra Help**, turn to **Appendix A**, **Topic 1**.

Extensions

Arnie's Sound

Arnie has a small business. He takes a master recording from a recording studio and transfers it to a compact disc. After a certain number of compact discs are manufactured, they go to the record company and are then sold to the public.

Arnie puts the material through four stages before returning the master recording and the CDs to the record company.

In the first stage, Arnie makes a master disc from the master recording. A laser cutter is used to cut pits into a master disc. The master disc is used to make the compact discs.

In the second stage, Arnie puts the master disc into one of the three stamping machines that his business owns. These machines stamp the master disc onto a clear plastic disc. This new disc holds the information that will reproduce the music in any compact disc player.

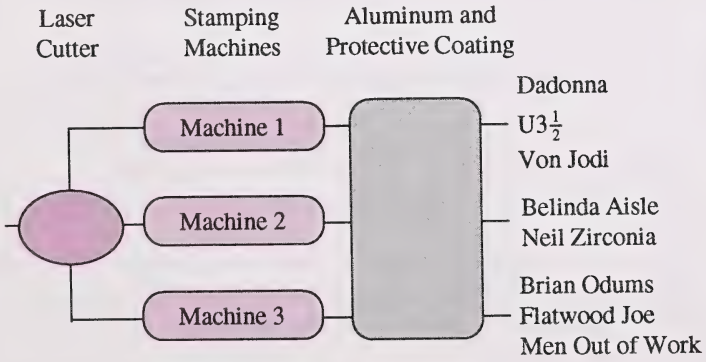
In the third stage, Arnie uses a machine that puts an aluminum coating on the clear plastic discs. The aluminum coating causes the light to be reflected back in the compact disc. The pits cut into the plastic disc will determine at what intensity the light will be reflected back. Following this, a protective clear plastic coating and the credits are put on the disc.

In the last stage, the discs are placed in their containers, which are made by a different company, and returned to the record company.

As you have probably already noted, there are two places in this process where Arnie will be testing the quality of his materials. The first check occurs when the master disc is made. At this point, every master disc will be checked. If a master disc is not properly made, none of the products will be properly made. Arnie will check quality again when the final protective coating and credits have been placed on the disc. At this point he will use the random number tables to randomly select the discs to be checked.

Since Arnie is only making one master disc for each artist, which is only placed on one of the stamping machines, he will randomly select discs for each artist as they are produced. This will allow him to automatically select discs randomly from all three of his stamping machines and from all of the people he has operating them.

Each of the artist's discs goes through Arnie's company as follows:



Using this information and the information you were given in Activity 2, do the questions that follow.

Do all of the following questions.

1. How many CDs will each stamping machine produce?

2. Arnie expects to spend $62\frac{1}{2}$ days stamping the CDs. If he has two shifts of eight working hours each, how many CDs can he stamp in one hour?

3. How many CDs must be checked per hour per machine if 3% of the CDs are to be checked?

4. How many CDs must be checked per machine and per artist if 1% of the CDs are to be checked?

5. How many CDs must be checked per hour per machine if 1% of the CDs are to be checked?

For solutions to **Extensions**, turn to **Appendix A, Topic 1.**

Topic 2 Organization and Presentation of Data

Introduction

After collecting data, such as information about which type of school dance to have, it is usually necessary to organize the data. This is done either to help with the decision that needs to be made, or to show someone else what was found. There are many ways to organize the data.

What Lies Ahead

Throughout the topic you will learn to

1. organize data using tables, stem-and-leaf plots, and box plots
2. graph data and interpret data graphed on pictographs, bar graphs, and circle graphs
3. graph and interpret continuous data using histograms and frequency polygons

Now that you know what to expect, turn the page to begin your study of organization and presentation of data.

Exploring Topic 2

Activity 1

Organize data using tables, stem-and-leaf plots, and box plots.

Once you have collected the data, it is necessary to organize it in a form which will clearly show any patterns or trends that might exist.

Tables

A table is probably the method with which you are most familiar. A table allows you to put data in a concise and easily read form while occupying a small area. Take another look at the tables that were used in the first topic.

The Size of Samples Used by Arnie's Sound

Size of Population	10	100	1000	10 000	100 000	1 000 000	10 000 000
Size of Sample	10	100	750	5000	10 000	50 000	200 000
Percentage of Population	100%	100%	75%	50%	10%	5%	2%

This table is designed to make a comparison between the size of a sample and a population as the size of the population increases. Here, the sample is becoming a smaller and smaller proportion of the population as the population increases.

The next table also shows a comparison between the size of the sample and the population. However, in this table, the actual size of the sample and the population are shown. This table shows that the proportion of the sample to the population remains the same so that the total sample is representative of the population.

Actual Sample Sizes Used For Testing by Arnie's Sound

Artist	Size of Population	Size of Sample	Percentage of Population
Belinda Aisle	300 000	9000	3%
Brian Odums	150 000	4500	3%
Dadonna	400 000	12 000	3%
Flatwood Joe	250 000	7500	3%
Men Out of Work	125 000	3750	3%
Neil Zirconia	175 000	5250	3%
U3$\frac{1}{2}$	750 000	22 500	3%
Von Jodi	350 000	10 500	3%
TOTAL	2 500 000	75 000	3%

Other tables can show how items make up a whole, or how frequently an item or score may occur.

Make a table using the information given in the previous table to show how the artists' CDs make up the total number of CDs manufactured by Arnie's Sound.

Percentage of Artists' CDs Manufacturered at Arnie's Sound

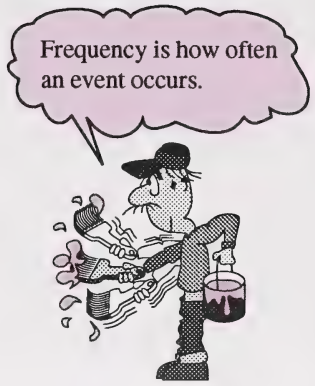
Artist	Size of Population	Percentage of Population
Belinda Aisle	300 000	12%
Brian Odums	150 000	6%
Dadonna	400 000	16%
Flatwood Joe	250 000	10%
Men Out of Work	125 000	5%
Neil Zirconia	175 000	7%
U3$\frac{1}{2}$	750 000	30%
Von Jodi	350 000	14%
Total	2 500 000	100%

Each percentage can be found by dividing the total size of the popluation into the size of the population for each artist, and then multiplying by 100. For example, the percentage of CDs manufactured at Arnie's Sound for Belinda Aisle is found as follows.

$$\frac{\text{Size of item}}{\text{Size of population}} \times 100 = \frac{300\,000}{2\,500\,000} \times 100$$
$$= 0.12 \times 100$$
$$= 12\%$$

Now look at a table where you will be keeping track of the **frequencies**.

The following table is used to keep track of the number of orders that had a specified percentage of CDs that failed the testing sequence.



Number of Batches with Percentage of Failed Discs

Percentage of Discs that Failed	Tally	Frequency
0.5	IIII	4
1.0	𝍸 II	7
1.5	𝍸 𝍸 III	13
2.0	𝍸 𝍸 IIII	14
2.5	𝍸 𝍸 I	11
3.0	𝍸 I	6
3.5	I	1

The percentage of failures has been rounded to the nearest 0.5%.

The following things can be read from this table.

- You can tell how many batches had a certain percentage of failed CDs. Four batches had 0.5% of the total number of discs that failed.
- You can tell how many CDs will normally fail from a given batch. You would expect that no more than 3.5% of the CDs will fail.
- How many batches are being used for this table? In this case you would add up the frequency column which gives you a total of 56 batches.

You will be using tables later for the construction of graphs.

Stem-and-Leaf Plots

This method has the advantages of both a table and a bar graph. Take a look at the following to see how it works.

In the month of December 1991, Arnie's Sound had the following number of failed discs from the 32 batches of discs that were manufactured.

60	58	25	87	93	55	24	70
62	49	58	64	59	60	**98**	55
66	57	48	12	67	66	58	52
58	78	38	44	**09**	61	83	52

Step 1: Find the lowest and highest values. The smallest number 9 has a 0 in the tens place, and the largest number 98 has a 9 in the tens place. Therefore, you choose the stems to be the digits 0 to 9 inclusively.

Write the stems vertically with a vertical line to the right of the column as shown.

Stem	Leaf
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	

Step 2: The ones digit in each number is a leaf. For each number in the table, write its leaf to the right of the stem. For the first number 60, write the digit 0 beside the number 6 as shown.

Stem	Leaf
0	
1	
2	
3	
4	
5	
6	0
7	
8	
9	

The process is continued until all of the digits have been entered as leaves.

Stem	Leaf
0	9
1	**2** ← This represents the number 12.
2	5 4
3	8
4	9 8 4
5	8 8 7 8 9 5 8 5 2 2
6	0 2 6 4 **7** 0 6 1
7	8 0 ↑
8	7 3 This represents the number 67.
9	3 8

Step 3: Arrange the leaves in ascending order (from smallest to largest). The resulting stem-and-leaf plot is given as follows.

The Number of Failed CDs
for 32 Batches at Arnie's Sound
During December, 1991

Stem	Leaf
0	9
1	2
2	4 5
3	8
4	4 8 9
5	2 2 5 5 7 8 8 8 8 9
6	0 0 1 2 4 6 6 7
7	0 8
8	3 7
9	3 8

With the diagram titled, it is now completed. As you can see from the previous plot, it has the advantages of having all of the values present and it has the general shape of a bar graph.

Stem	Leaf
0	9
1	2
2	4 5
3	8
4	**4 8 9**
5	2 2 5 5 7 8 8 8 8 9
6	0 0 1 2 4 6 6 7
7	0 8
8	3 7
9	3 8

In the previous steam-and-leaf plot, **4** | **4 8 9** could be thought of as follows:

The 4 is the branch or stem, and the digits 4, 8, and 9 are the leaves.

In the example that you have just gone through, the stem is in the tens place and the leaves are in the ones place. This is not always true. Both the stem and the leaves can be in any given place value. The only rule is that the stem will always be a larger place value than the leaves.

Another example of stem-and-leaf place values would be the hundreds place for the stem and the tens and ones digits for the leaves.

Here is an example where the hundreds digit is the stem and the tens and one digits are the leaves.

In the six months ending on December 31, 1991, Arnie's Sound has the following number of failed discs from the batches that they manufactured.

603	582	256	872	935	554	249	702
623	494	585	642	592	609	983	556
664	572	482	129	673	663	582	523
585	783	384	442	97	618	834	526

This data could be displayed in a stem-and-leaf diagram in this way.

Stem	Leaf
0	97
1	29
2	56 49
3	84
4	94 82 42
5	85 82 72 85 92 54 82 56 23 26
6	03 23 64 42 73 09 63 18
7	83 02
8	72 34
9	35 83

All that remains is to arrange the leaves in ascending order.

Stem	Leaf
0	97
1	29
2	49 **56** ← This represents the number 256.
3	84
4	42 82 94
5	23 26 54 56 72 82 82 85 85 92
6	03 09 18 23 42 63 64 73
7	02 83
8	34 72
9	**35** 83

↑
This represents the number 935.

The tens and the ones digits can be regarded as ones since any number of tens can be changed to ones.

For example,

$623 = 600 + 23$

$= 600 + 20 + 3$

or $= 600 + 23$ ones

It is for this reason that the last two digits are sometimes referred to as ones.

Box-and-Whisker Plots

Box-and-whisker plots are used to show how data is distributed or spread out. You will use the same data to develop a box-and-whisker plot as was used for the stem-and-leaf plot.

In the six months ending on December 31, 1991, Arnie's Sound had the following number of failed discs from the 32 batches of discs that they manufactured.

603	582	256	872	935	554	249	702
623	494	585	642	592	609	983	556
664	572	482	129	673	663	582	523
585	783	384	442	97	618	834	526

Step 1: Rearrange the data so that it is in numerical order, beginning with the smallest value and ending with the largest value.

97	129	249	256	384	442	482	494
523	526	554	556	572	582	582	585
585	592	603	609	618	623	642	663
664	673	702	783	834	872	935	983

Step 2: Find the median of the arranged data.

- The median is the middle number of a group of numbers that have been arranged in ascending or descending order.
- The median of 8, 5, **4**, 2, 1, is 4.
- The median of an even number of data is found by taking the average of the two middle numbers.
- The median of 8, **5**, **4**, 2, is $\frac{5+4}{2} = 4.5$.

97	129	249	256	384	442	482	494
523	526	554	556	572	582	582	**585**
585	592	603	609	618	623	642	663
664	673	702	783	834	872	935	983

The data is broken down into two equal halves. The median will be the average of the last number in the first group and the first number in the second group. Since both numbers are 585, the median will be 585.

Step 3: Find the median of the upper half and the lower half.

97	129	249	256	384	442	482	**494**
523	526	554	556	572	582	582	585

The median of the first group will be $\frac{494+523}{2} = 508.5$.

585	592	603	609	618	623	642	**663**
664	673	702	783	834	872	935	983

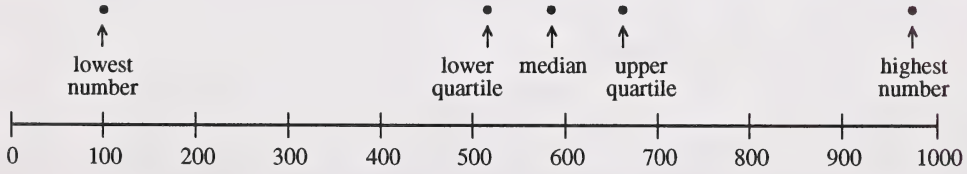
The median of the second group will be $\frac{663+664}{2} = 633.5$.

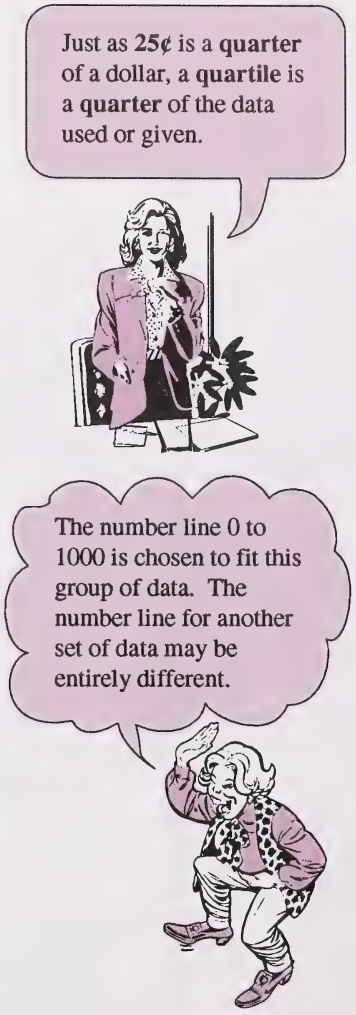
The three points which divide a set of data into four equal parts are called quartiles. The middle quartile is the **median** and the other two are the **lower** and **upper** quartiles.

Step 4: Note the lowest and highest values. The lowest value is 97 and the highest value is 983.

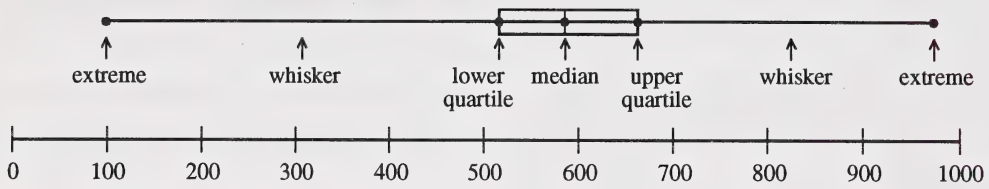
All of the actual calculating is finished. Now comes the time for drawing. Draw a number line that goes slightly lower than the lowest value and slightly higher than the highest value.

Step 5: Mark dots above the line, as shown, for the three medians, the highest number, the lowest number.



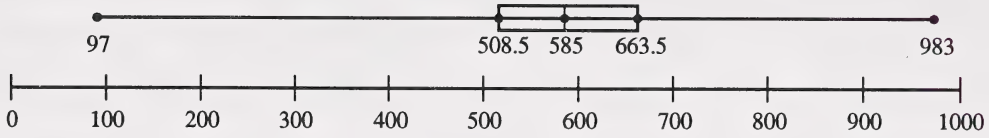

Just as 25¢ is a quarter of a dollar, a quartile is a quarter of the data used or given.
The number line 0 to 1000 is chosen to fit this group of data. The number line for another set of data may be entirely different.

Step 6: Draw a box using the lower quartile and the upper quartile, and draw whiskers to the extremes. Draw a line across the box to mark the median.

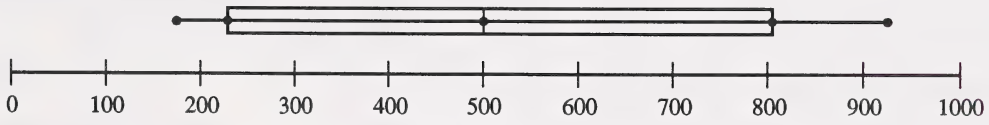


Step 7: Put the appropriate title onto the plot.

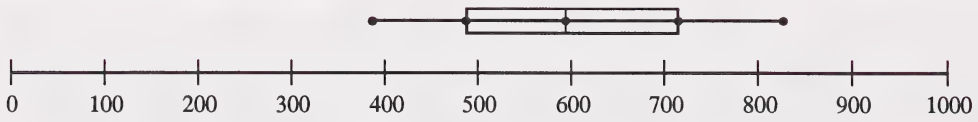
The Number of Failed Discs From 32 Batches at Arnie's Sound
July 1 to December 31



This box-and-whisker plot has a small box and long whiskers. Not all box-and-whisker plots will be the same. This one shows that the data is stretched out over a large number of values, but half of the data is concentrated in a small area, which is the box. One-half of the data is located between the lower quartile 508.5, and the upper quartile 663.5, which is represented by the rectangle.



This box-and-whisker plot is a different case. The box and whiskers are stretched over a large number of values, but this time the box is large and the whiskers are short. The lower extreme is about 185, the upper extreme is about 935, the lower quartile is about 230, the upper quartile is about 805, and the median is about 500. This box-and-whisker plot shows that the data is more concentrated at the extremes.



In this case, the box and the two whiskers occupy approximately the same space. Thus, the values are evenly distributed.

Try the following questions.

Do questions 1, 3, 5, 6, and 7 **or** questions 2, 4, 5, 6, and 7.

1. An oil company found oil at the following depths (in metres) on one of their oil fields.

378	311	352	344	359	327	362	382
316	364	352	385	309	361	363	356
383	364	334	338	364	352	357	354

   a. Organize this data into a stem-and-leaf plot.

   b. Organize this data into a box-and-whisker plot.

2. The populations of a group of towns in Alberta are given as follows.

1024	1267	1876	1604	1554	1143	1730	1840
1920	1266	1544	1739	1548	1732	1434	1498
1603	1436	1288	1356	1497	1098	1872	1064

   a. Organize this data into a stem-and-leaf plot.

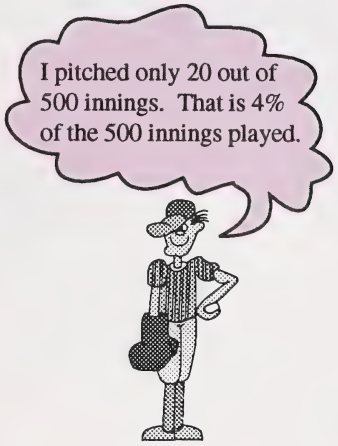
   b. Organize this data into a box-and-whisker plot.

3. The following is the number of items sold of each type of food in a fast food restaurant. Make a table showing what percentage each item represents of the total amount sold that day. Round your answers to the nearest whole percent.

Hamburgers	134	Onion rings	29
Hot dogs	26	Pop	132
French fries	98	Coffee	81

4. The number of innings pitched by a pitching staff is given as follows. Arrange the data in a table that will show the percentage of innings that each player pitched as compared to the total number of innings played. Round your answers to the nearest whole percent.

Dodd	90	Freelance	123	Garcia	67
Foggson	72	Inrate	23	Morris	90
Stanton	12	Worson	20	Zielert	3



5. Use the following table to answer the questions that follows.

How Different Types of Music Sells
(Survey made in the A&S Record Store on February 2, 1992)

Type of Music	Frequency	Percentage of Total
Blues	12	6%
Classical	32	16%
Country	34	17%
Reggae	6	3%
Soul	42	21%
Rock	74	37%
Total	200	100%

a. What type of music is the most popular?

b. What type of music is the least popular?

c. How many people were talked to in the survey?

d. What percentage of the people prefer country music?

e. Where and when was this survey done?

f. Do more people prefer rock music or blues and soul music combined? Support your answer.

6. Use the following stem-and-leaf plot to answer the questions that follow.

The Number of Home Runs Hit by the Bulls
During the 1992 Season After Playing 120 Games

Stem	Leaf
0	0 2 3 8 9
1	0 1 1 5 7 7 8
2	0 2 4 4 7 8 8 9
3	2 5 6 8
4	3
5	2

a. How many players are on the Bulls team?

b. How many players hit more than twenty-five home runs?

c. How many games did the Bulls play?

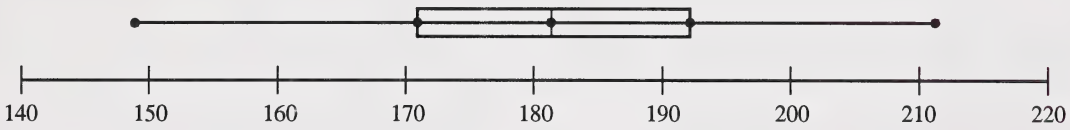
d. What was the largest number of home runs hit by one player?

e. What was the least number of home runs hit by one player?

f. What are the exact number of home runs hit by each player that hit between 30 and 39 home runs?

7. Use the following box-and-whisker plot to answer the questions that follow.

The Height (in cm) of Students in Captain Gerald Riesman High School



a. What is the height of the shortest student in the school?

b. What is the height of the tallest student in the school?

c. What is the median height of the students in the school?

d. Between what two measures will you find half of the students' heights? More than one answer is possible.

e. What unit of measurement is being used?

f. Describe how the heights of the students are distributed.

For solutions to **Activity 1**, turn to **Appendix A**, **Topic 2**.

Activity 2

Graph data and interpret data graphed on pictographs, bar graphs, and circle graphs.

Once the data has been collected and organized, it is ready to be used in making a graph.

What is a graph? A graph is a method displaying data that will permit you to compare different aspects of the data with only a glance.

Why not use a graph to display all data instead of tables and plots? What is gained in the speed of comparing the data in a graph is lost in the accuracy of the data being displayed.

The numbers collected from the data are only being estimated when the visual symbols of the graphs are used. In the bar graph, the length of the bar must be estimated. For the pictograph, the proportion of a picture that is being used must be estimated. For the circle graph, you have to estimate what the actual figures are from the percentages given on the graph.

Take a closer look at each type of graph and see how to choose the type of graph you should use, and how to read information from that graph.

The Pictograph

The pictograph is identified by the use of small pictures to represent a specified amount of data. Examine the following pictograph and see what makes up a proper pictograph.

Canuckland's Five Largest Industries
Its Total Assets in 1992

BBE	$$$$$ $$$$$ $$$$$ $$$$$ $$$$$ $$
Canuckland International	$$$$$ $$$$$ $$$$$ $$$
Canuckland Government Systems	$$$$$ $$$$$ $$$$$ $$$$$ $$$$$ $$$$$ $$$
Centreland Hydro	$$$$$ $$$$$ $$$$$ $$$$$ $$$$$ $$$$$ $$
BMS	$$$$$ $$$$$

Legend: $ = 1 000 000 000 dollars
Source: 1992 Canuckland World Almanac

The single factor that identifies this graph as a pictograph is the dollar sign $. In this graph, the $ represents one billion dollars of assets.

Characteristics of a Pictograph

- A title to inform the reader what is being shown in the graph must be identified.
- The items or categories must be clearly identified. These are usually written on the left side of the graph. However, it is now possible to find graphs in magazines that do not use the left margins. In these cases, the magazines are trying to make the graph more interesting to the reader.
- The pictures must appear next to the category labels. In this particular graph, the $ signs are grouped in fives. This makes it easier for the reader to count the amounts.
- A legend must be present to tell the reader what each picture represents.

A pictograph is used when you want to make a visual comparison of similar data. The data is grouped into separate and distinct categories.

Take a look at how to construct a pictograph. The following table shows the size of the populations of CDs made by Arnie's Sound.

Step 1: Select an appropriate title for the pictograph.

Since a title is already given in the table, use it as the title of the pictograph.

Actual Population Sizes Used
for Testing by Arnie's Sound

Artist	Size of Population
Belinda Aisle	300 000
Brain Odums	150 000
Dadonna	400 000
Flatwood Joe	250 000
Men Out of Work	125 000
Neil Zirconia	175 000
U3$\frac{1}{2}$	750 000
Von Jodi	350 000
Total	2 500 000

Step 2: Put the category labels on the graph.

In this case, the categories will be the names of the artists. They will be written down the left margin.

Actual Population Sizes Used for Testing by Arnie's Sound
Belinda Aisle
Brain Odums
Dadonna
Flatwood Joe
Men Out of Work
Neil Zirconia
U3$\frac{1}{2}$
Von Jodi
Total

Step 3: Select the symbol to be used as a picture in the graph. Since compact discs are used, use the following symbol.

Step 4: Decide on what the symbol will represent and write this down on the graph as a legend.

For this particular graph, let the symbol represent 50 000 CDs.

Actual Population Sizes Used for Testing by Arnie's Sound
Belinda Aisle
Brain Odums
Dadonna
Flatwood Joe
Men Out of Work
Neil Zirconia
U3$\frac{1}{2}$
Von Jodi
Legend: = 50 000 CDs

The legend must be chosen so that all of the data can be displayed in a reasonable amount of space.

If you let each symbol represent 100 discs, then there would not be enough space to put in the 3000 symbols required to represent the 300 000 Belinda Aisle discs sold.

Likewise if you let each symbol represent 1 000 000 discs, only parts of a disc would be used for each category. This would make it very difficult to read the number of discs made for each artist.

Step 5: You must decide how many pictures will be needed for each category. To do this, divide the number each symbol represents into the number of discs for each category.

The following are the calculations used to determine the number of symbols to be used for each category.

Belinda Aisle: $\frac{300\,000}{50\,000} = 6$

Brian Odums: $\frac{150\,000}{50\,000} = 3$

Dadonna: $\frac{400\,000}{50\,000} = 8$

Flatwood Joe: $\frac{250\,000}{50\,000} = 5$

Men Out of Work: $\frac{125\,000}{50\,000} = 2.5$

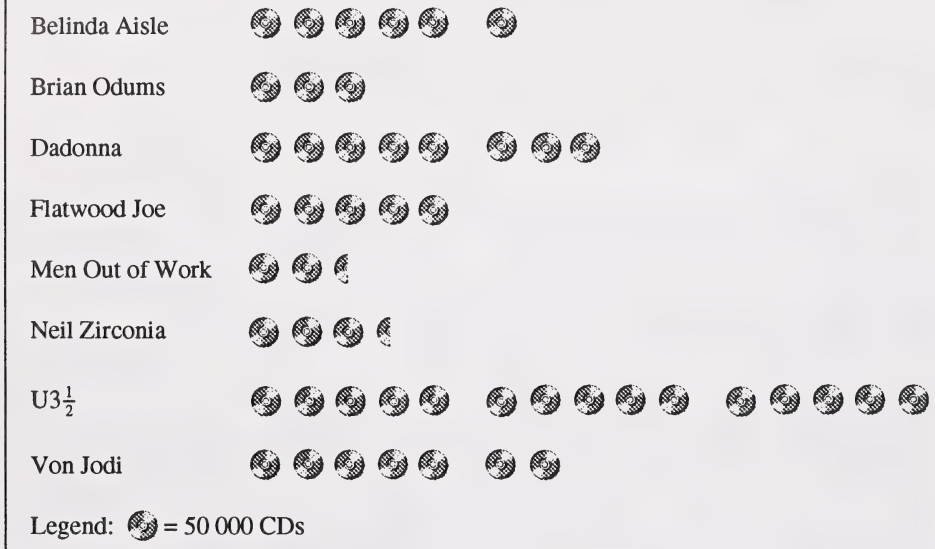
Neil Zirconia: $\frac{175\,000}{50\,000} = 3.5$

U3$\frac{1}{2}$: $\frac{750\,000}{50\,000} = 15$

Von Jodi: $\frac{350\,000}{50\,000} = 7$

Step 6: Draw in the pictures. Keep the pictures in a straight line.

Actual Population Sizes Used for Testing by Arnie's Sound



Notice that half of a symbol was used to represent 25 000 CDs. One-half of 50 000 CDs or 25 000 CDs is represented by .

Other features such as inserting lines is to help separate parts of the graph that are not necessary. You may choose to insert them as they will make your graph easier to read.

The graph is now complete.

Steps for making a pictograph

- Select an appropriate title.
- Label the categories.
- Select an appropriate symbol.
- Select an appropriate scale.
- Calculate the number of symbols needed for each category.
- Draw the pictures on the graph.

Not only do you have to be able to make a pictograph, you must also be able to read one.

See if you can answer the following questions from the pictograph.

- Which artist had the most CDs made?

  Look at the pictograph and find the category that has the most discs beside it. In this particular case, it is U3$\frac{1}{2}$.

- How many CDs were made of Dadonna's music?

  Count the number of discs to the right of Dadonna; then, multiply that number by 50 000. The 50 000 is the scale size.

  $8 \times 50\,000 = 400\,000$

  There were 400 000 CDs made of Dadonna's music.

- Which artist's CD had 350 000 copies made?

  Divide the 350 000 by 50 000 to find out how many discs should be shown in the category.

  $350\,000 \div 50\,000 = 7$

  Look for the category that has seven CDs. In this case, it would be Von Jodi.

- Who made the artist' CDs?

  Look at the title of the pictograph to identify the manufacturer. In this case it is Arnie's Sound.

Take a look at another type of graph that is very similar to the pictograph. It is called the bar graph.

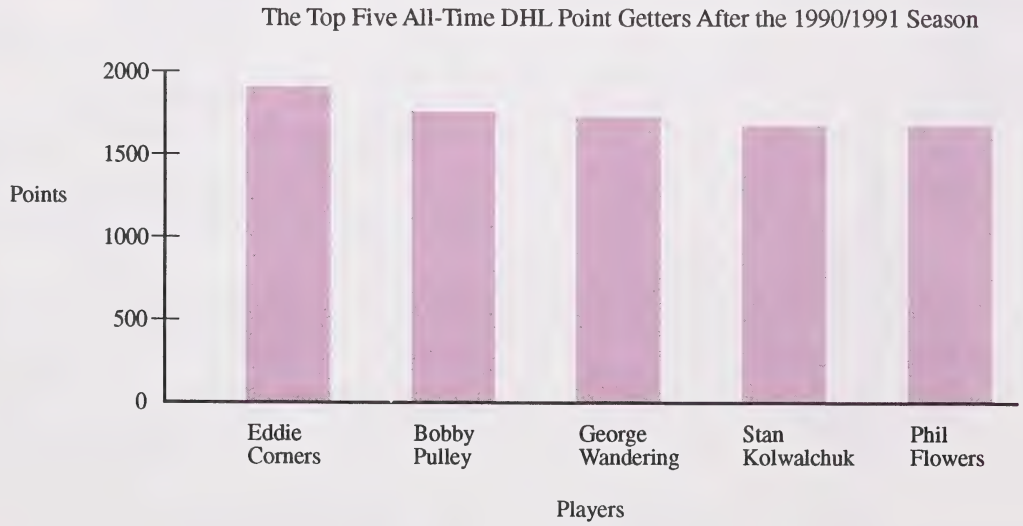
The Bar Graph

There are only three differences between the bar graph and the pictograph: the pictures are replaced by a bar, the categories are usually listed along the bottom, and the legend is replaced by a scale along the left side of the graph.

The length of the bar represents the numerical value of the data. To estimate the value of each bar, find the top of the bar and move horizontally over to the scale on the left side of the graph to read the value. All of the bars on the graph should have the same width and should have equal spaces between them.

The scale along the left side of the graph should always start at zero and move up in equal measurements.

Take a look at a bar graph.



All bar graphs have the following features.

- A title to inform the reader what is being shown in the graph.
- Items or categories that are clearly labelled. They are usually written along the bottom of the graph.
- The bars of the graph must be above or beside its corresponding item label, must be of equal width, and must have equal distances between them.
- The axes must be labelled with numerical values and a label, must increase by **equal** values from 0, and must be present on at least one side of the graph.

A bar graph is used when you want to make a visual comparison of similar data. The data is grouped into separate and distinct categories.

This is the same situation where the pictograph is used. How do you decide which of the two graphs to use?

The bar graph gives a more accurate estimate of the numerical values while the pictograph usually has a more appealing appearance.

Now look at how to construct a bar graph. You will use the same table that was used to make the pictograph.

Actual Population Sizes Used
for Testing by Arnie's Sound

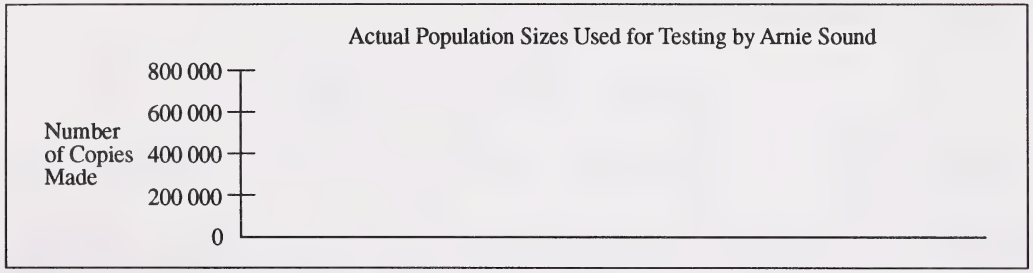
Artist	Size of Population
Belinda Aisle	300 000
Brain Odums	150 000
Dadonna	400 000
Flatwood Joe	250 000
Men Out of Work	125 000
Neil Zirconia	175 000
U3$\frac{1}{2}$	750 000
Von Jodi	350 000
Total	2 500 000

Step 1: Select an appropriate title for the bar graph.

A title is already given in the table. Use the same title for the bar graph.

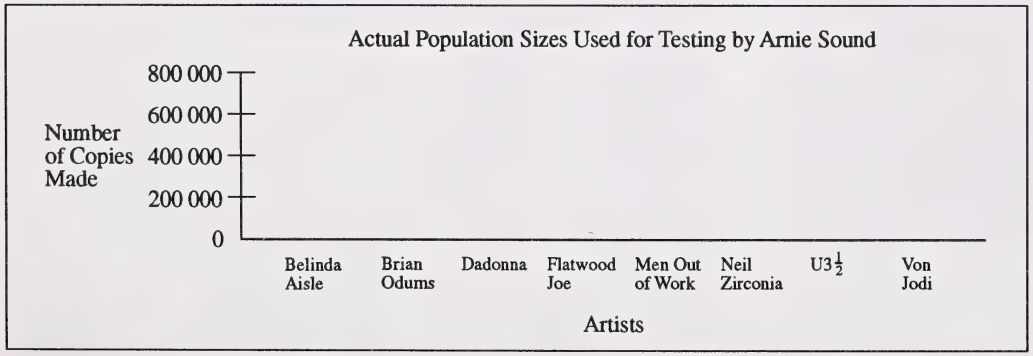
Step 2: Select and label an appropriate scale for the graph along the vertical axis. The scale must start at zero.

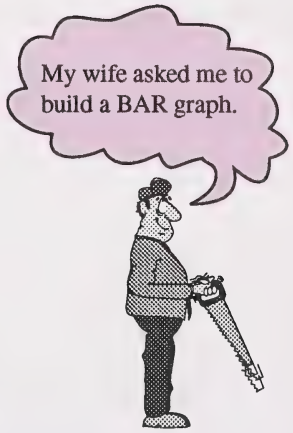
The scale should be selected so that the tallest bar will not quite reach the top of the graph. For this particular case, the tallest bar will reach 750 000; therefore, number the scale up to 800 000.



Step 3: Put the categories and the labels along the horizontal axis of the graph.

In this particular case, the categories will be the names of the artists. Write them along the bottom of the graph.

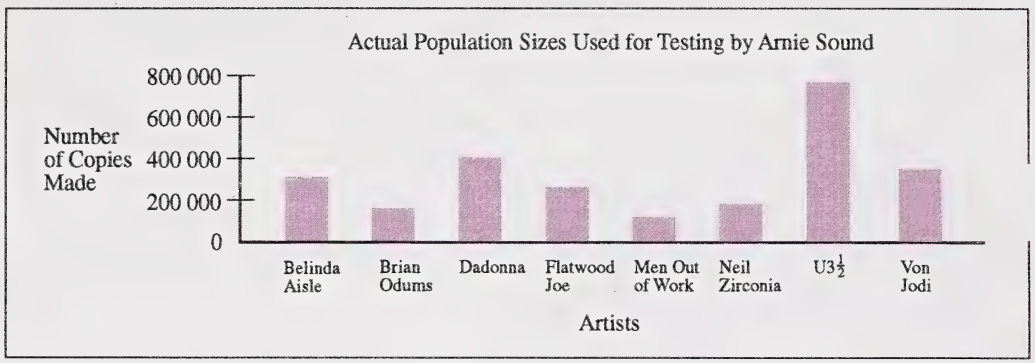




Step for constructing a bar graph.

- Select an appropriate title.
- Select an appropriate scale.
- Label the categories.
- Draw the bars.

Step 4: Draw the bars making sure that each bar is the correct length. Make sure all of the bars have the same width and that the space between the bars is the same.



If you have trouble getting equal space between the bars and getting them centered over the items, try using graph paper. Be careful that you get the proper item labels underneath the bars. You may color the bars to make the graph more interesting and more appealing.

The graph is now complete.

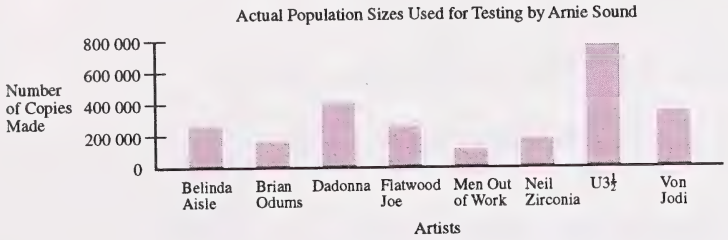
Do you know how to read information from this graph? See if you can find the answers to the following questions.

- Which artist's CD had the least copies made?

  Look for the shortest bar on the graph. On this particular graph, it is the bar that represents Men Out of Work.

- How many CDs were made of Flatwood Joe's Music?

  Go to the top of the bar that represents Flatwood Joe. Then go horizontally across the graph, to the left, and read off the value along the axis. The value is about 250 000 CDs.



- Which artist's disc had 175 000 copies made?

  For this question, start from the scales on the left. Find the 175 000 position on the scales. Then move horizontally to the right across the graph until you come to the top of one of the bars. You will come to the top of the bar that represents Neil Zirconia.

- What do these bars represents?

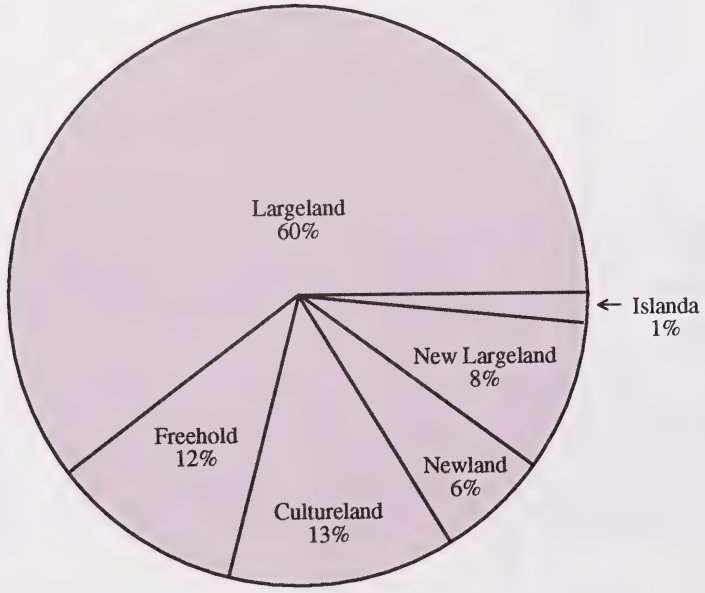
  The answer to this question will be found in the title. According to the title, each bar represents the population of the CDs made of each artist's recordings, or the number of copies of each artitst's disc.

Now take a look at the circle graph.

The Circle Graph

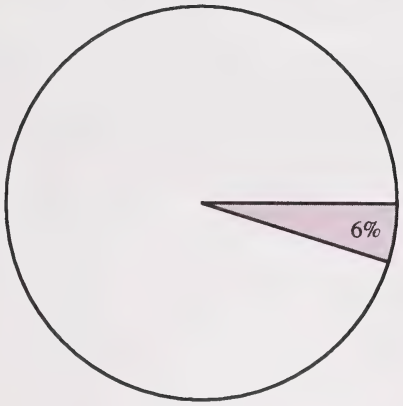
The circle graph is identified by its circular shape. The circle is divided into pie-shaped pieces that represent the different parts of the whole. Examine the following circle graph and see what characteristics make up a proper circle graph.

The Distribution of the Population of Canuckland
(The population is 32 million.)



Circle graphs have the following features.

- A title to informs the reader what is being shown in the graph.
- The items or categories must be clearly identified. The label is usually written in the piece of the circle that it represents. When there is not enough space for the label it is written beside the piece.



- Each label usually includes a percentage to identify the size of the piece. The percentages usually add up to 100%. It may also include the actual numerical value that the piece represents. For example, Largeland might have included 19.2 million inside the piece.

  To find the population of Largeland, use the following calculation.

  60% of 32 million

  $= 0.60 \times 32\,000\,000$

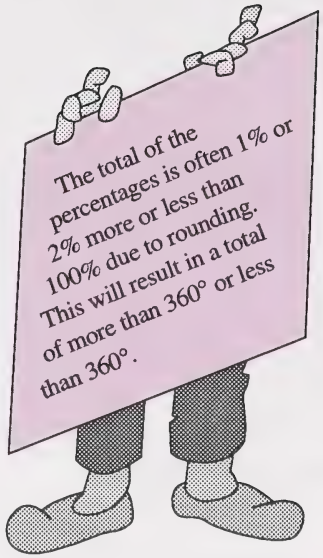
  $= 19\,200\,000$

  or 19.2 million

- The total numerical value of all the pieces should be included with the graph.

A circle graph is used when you want to compare the amounts of all separate items or parts to a whole.

The circle is divided into portions each representing a percentage of the whole. This percentage is used to label the portion of the graph. For example, the portion for Largeland on the graph occupies 60% of the whole circle, the same percentage that is stated on the piece.


The total of the percentages is often 1% or 2% more or less than 100% due to rounding. This will result in a total of more than 360° or less than 360°.

Take a look at how to construct a circle graph. You will use the same data that was used for the pictograph and the bar graph. This data is in the table that follows.

Actual Population Sizes Used for Testing by Arnie's Sound

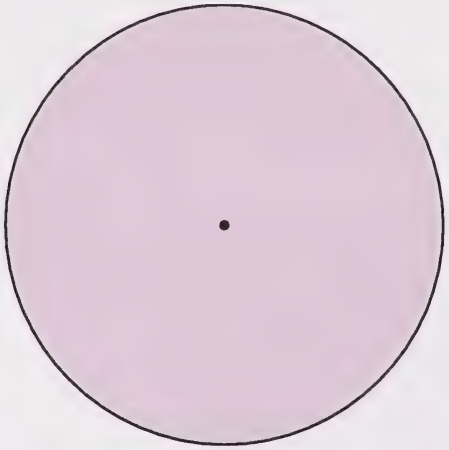
Artist	Size of Population	Population
Belinda Aisle	300 000	12%
Brain Odums	150 000	6%
Dadonna	400 000	16%
Flatwood Joe	250 000	10%
Men Out of Work	125 000	5%
Neil Zirconia	175 000	7%
U3$\frac{1}{2}$	750 000	30%
Von Jodi	350 000	14%
Total	2 500 000	100%

Step 1: Select an appropriate title for the graph.

Use the title that is used to name the table for the circle graph.

Step 2: Construct a circle. Be sure to identify the centre of the circle using a distinct dot.

Percentage of Artist's CDs Manufactured at Arnie's Sound



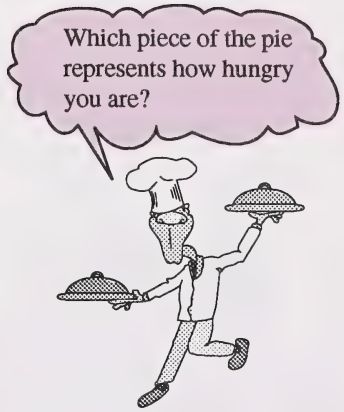
Step 3: Decide what percentage of the area each item or category will occupy.

This is done by dividing the amount of each item by the total amount, and multiplying by 100.

For example, to find the percentage for Belinda Aisle, divide 300 000 by 2 500 000. Then multiply by 100.

$$\frac{300\,000}{2\,500\,000} \times 100 = 0.12 \times 100$$

$$= 12\%$$



Fortunately for you, this has already been done on the table that you are using.

Step 4: Change the percentage to an angle measure. The measure of the angle will be used to determine the size of each piece on the graph.

A whole circle contains 360°. You will use the percentage to calculate the measure of the angle for each piece. The number of Belinda Aisle discs that were manufactured was 300 000. This is 12% of the total discs manufactured. Thus, the piece that would represent Belinda Aisle discs would be 12% of 360°.

Belinda Aisle

$$12\% \text{ of } 360° = \frac{12}{100} \times 360°$$
$$= \frac{4320}{100}$$
$$= 43.2$$
$$\doteq 43\%$$

Brian Odums

$$6\% \text{ of } 360° = \frac{6}{100} \times 360°$$
$$= \frac{2160}{100}$$
$$= 21.6$$
$$\doteq 22\%$$

Dadonna

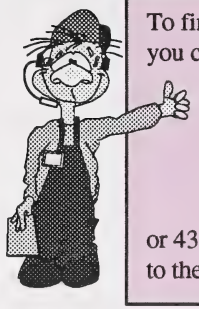
$$16\% \text{ of } 360° = \frac{16}{100} \times 360°$$
$$= \frac{5760}{100}$$
$$= 57.6$$
$$\doteq 58\%$$

Flatwood Joe

$$10\% \text{ of } 360° = \frac{10}{100} \times 360°$$
$$= \frac{3600}{100}$$
$$= 36°$$

Men Out of Work

$$5\% \text{ of } 360° = \frac{5}{100} \times 360°$$
$$= \frac{1800}{100}$$
$$= 18°$$



To find 12% of 360° you can also do this.

$$12\% \text{ of } 360°$$
$$= 0.12 \times 360°$$
$$= 43.2°$$

or 43° when rounded to the nearest degree.

Neil Zirconia

$$7\% \text{ of } 360° = \frac{7}{100} \times 360°$$
$$= \frac{2520}{100}$$
$$= 25.2$$
$$\doteq 25°$$

U3$\frac{1}{2}$

$$30\% \text{ of } 360° = \frac{30}{100} \times 360°$$
$$= \frac{10\,800}{100}$$
$$= 108°$$

Von Jodi

$$14\% \text{ of } 360° = \frac{14}{100} \times 360°$$
$$= \frac{5040}{100}$$
$$= 50.4$$
$$\doteq 50°$$

Check your calculations by adding up all the degree measures. They add up to 360 degrees.

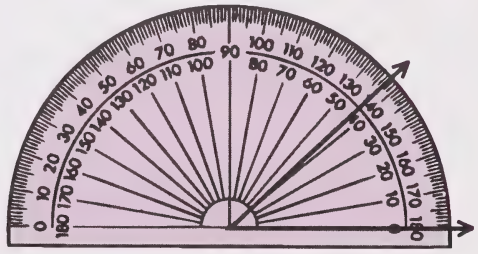
$$43° + 22° + 58° + 36° + 18° + 25° + 108° + 50° = 360°$$

The degreee measures add up to 360°. You do not need to go back and redo your calculations.

Step 5: Mark off the areas on your graph.

Using a protractor, measure the degrees around the centres of the circle. Use the following procedure.

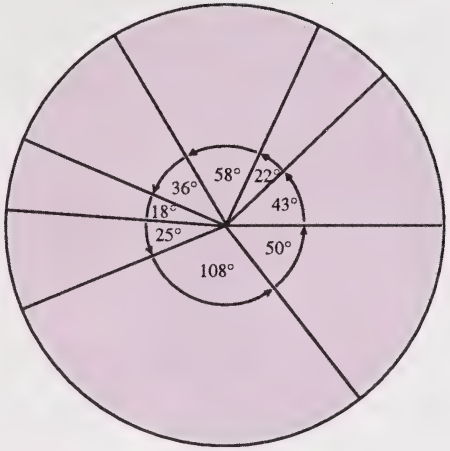
- Draw a line from the centre to the edge of the circle.
- Put the centre point of the protractor on the centre of the circle, and the base line along the line you have just drawn.



- Measure the angle and draw a line from the centre to the edge of the circle.

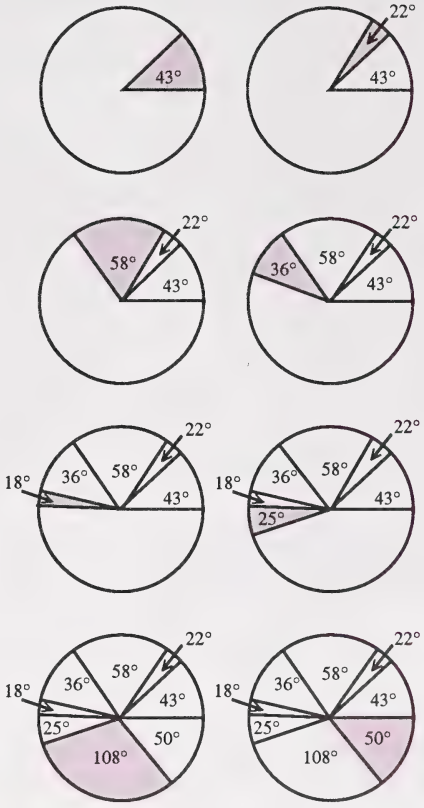
- Use this new line as your base line, and repeat this procedure until you have completed all of the pieces.

Percentage of Artist's CDs Manufactured at Arnie's Sound



* Begin here and follow the arrows. Here you progressed in a counterclockwise direction. You also could have proceeded in a clockwise direction.

The following diagrams show where to begin. Continue in order until all items are included. The diagrams also show how the number of degrees are used (if used) to label the various points.

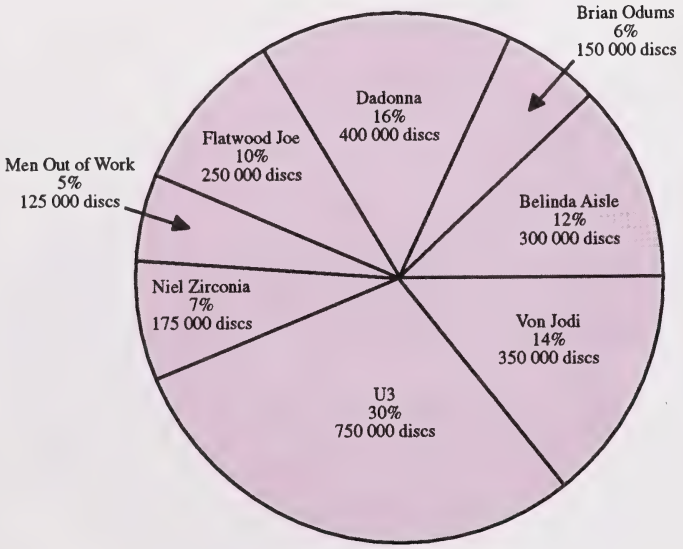


Step 6: Name the sections and put in the percentages for all the pieces. For this particular graph, show the numerical values that each piece represents.

Step 7: Indicate the numerical value on the graph that represents the whole. On the graph, it is placed just below the title.

The following is the completed graph.

Percentages of Artist's CDs Manufactured at Arnie's Sound



Do you know how to read the information presented on this type of graph? See if you can find the answers to the following questions.

- Which artist's disc had the most copies made?

  Look for the larget piece on the graph. This piece will represent the artist who has the most copies made. For this graph it would be U3$\frac{1}{2}$.

- How many copies were made of Von Jodi's music?

  Look for the piece of the circle that represents Von Jodi. Once this piece is found, you can read the information directly from the graph.

  If the actual number is not given, calculate by using the percent.

$$14\% \text{ of } 2\,500\,000 = \frac{14}{100} \times 2\,500\,000$$
$$= 350\,000$$

- Which artist's discs had 400 000 copies made?

  To find which artist's disc had 400 000 copies made, look for the piece of the graph that has a label of 400 000 discs.

  If the actual number is not shown, then calculate what percent 400 000 is of the total.

$$\frac{400\,000}{2\,500\,000} \times 100\% = 16\%$$

Now find the piece that is labelled 16%. In this case, it is Dadonna.

- What percentage of the total discs of Brian Odums' music were made?

  To find the percentage of discs made of a particular artitst's music, look for the artist's name on one of the pieces of the graph. Once you have found that particular piece, read the percentage that is in the label. For Brian Odums, this is 6%.

  If only the actual number is given, then calculate the percent as follows.

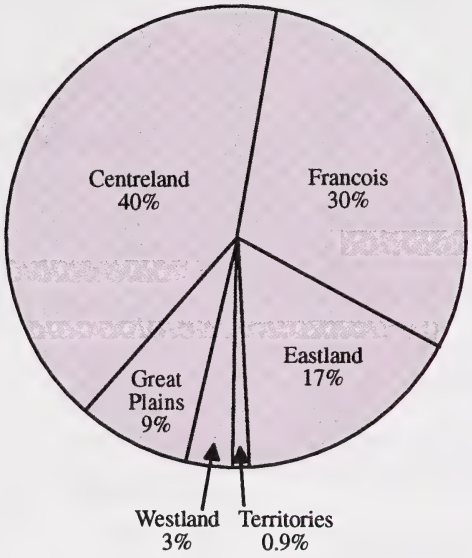
$$\frac{150\,000}{2\,500\,000} \times 100\% = 6\%$$

If you have access to a videocassette recorder, you may view the *Math Wise* video titled **Locating and Interpreting – Graphs.** This program discusses how to make and use graphs.

You should now be familiar with pictographs, bar graphs, and circle graph. Try the questions that follow.

1. Use these circle graphs to answer the questions about Adanacland. The questions are found after the graphs.

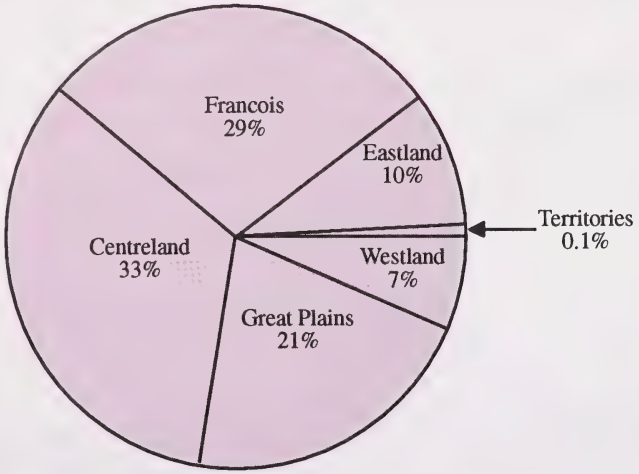
1901: Population 5 371 000



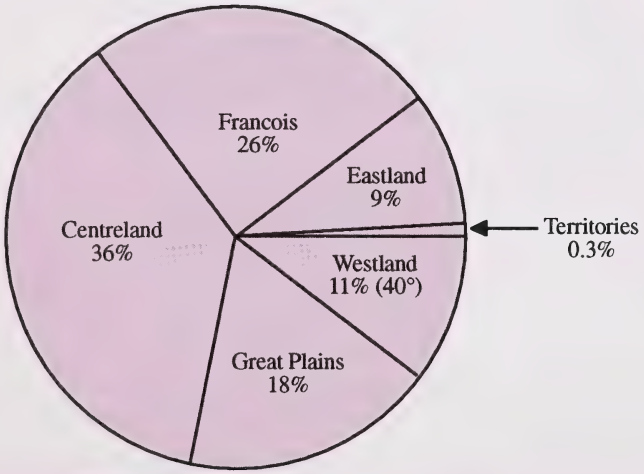
Step for constructing a circle graph.

- Select an appropriate title.
- Construct the circle that will be used for the graph.
- Find the percentage of each item.
- Change the percentage to an area measurement using degrees.
- Mark off the degrees.
- Name the labels.
- Indicate the units and the total numerical value that the graph represents.

1941: Population 11 507 000



1986: Population 25 384 000



a. Explain what information the graphs are illustrating.

b. Do the percentages for each graph total 100%?

c. In which areas of the country has the percentage of the total population decreased through all three graphs?

d. What is the population of Centreland in each of the graphs? Round your answers to the nearest thousand.

e. Which part of the graph decreased the most from the 1901 graph to the 1986 graph?

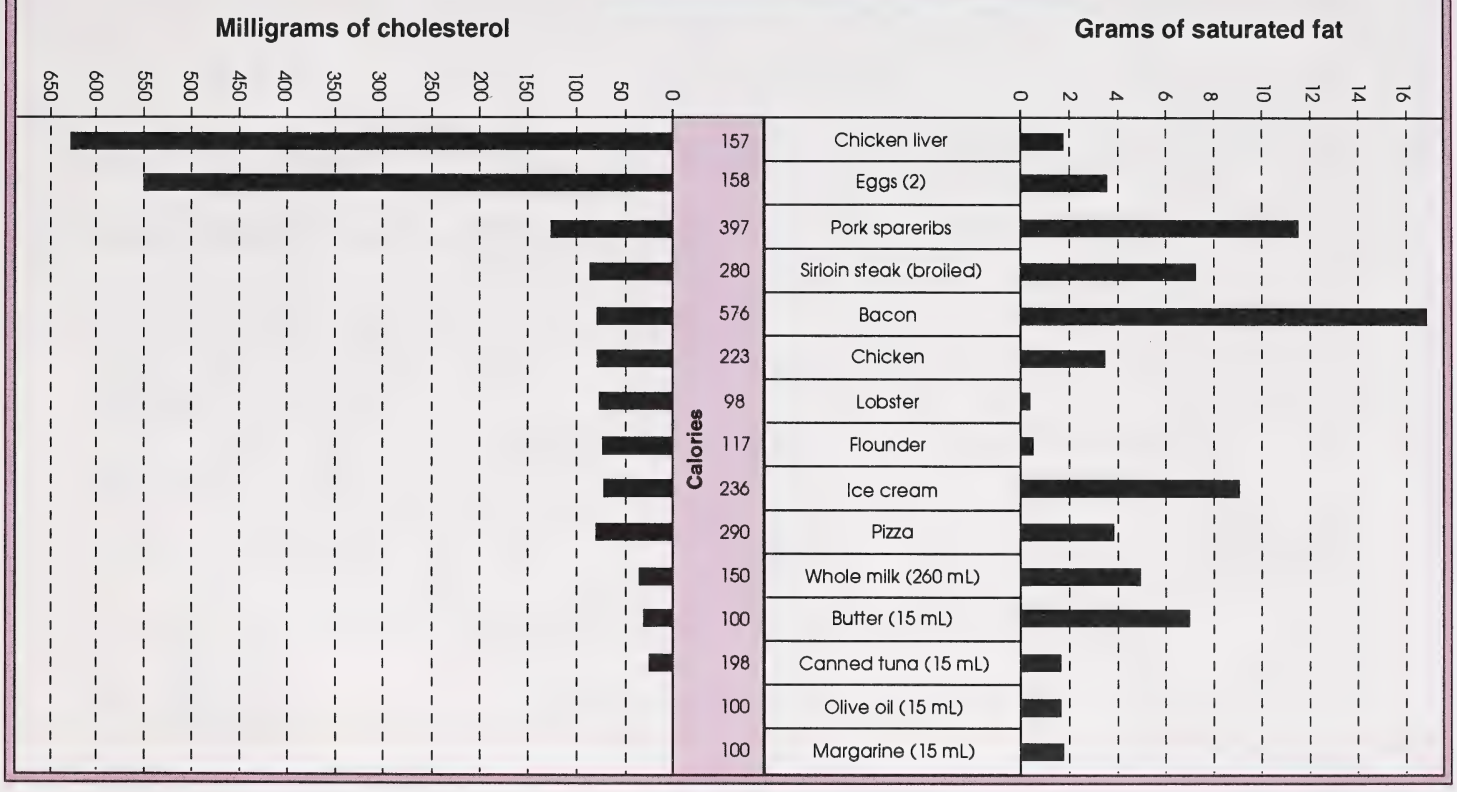
f. What was the total population for Adanacland in 1901, in 1941, and in 1986?

g. Which area of the country had the greatest percentage growth in population from 1901 to 1986?

h. The percentage of the total population for Great Plains has doubled from 1901 to 1986. Has its actual population doubled?

FACTS ON FAT

105 grams of each food, except where noted



2. Use the graph on the previous page to answer the following questions.

   a. Different quantities were used for the various foods. What were all of these different quantities?

   b. Which food has the least amount of cholesterol? What is that amount?

   c. Which food has the least amount of saturated fat? What is that amount?

   d. Which food has 223 calories?

   e. What two scales are being used in this graph?

   f. Which food has the highest amount in each of the categories?

   g. How many milligrams of cholesterol are there in 1L of milk? Keep in mind that $1L = 1000\ mL$.

   h. Which food has the highest level in two of the categories?

3. Use this graph to answer the following questions.

The Number of Employees for Canuckland's Seven Largest Employers

Each symbol represents 10 000 people.

   a. How many people does each symbol represent?

   b. How many people work for the Eddie Watson Company?

   c. How many more people work for CPS than Southern Fried Fish?

   d. How many employees are represented on this graph?

4. Make a pictograph to represent the number of automobile owners in Canada in 1991.

Newfoundland	176 000	Prince Edward Island	56 000
Nova Scotia	337 000	New Brunswick	286 000
Quebec	2 614 000	Ontario	4 244 000
Manitoba	527 000	Saskatchewan	389 000
Alberta	1 296 000	British Columbia	1 527 000
Territories	56 000		

5. Make a bar graph to represent the longest rivers in Canada. Source: Statistics Canada: Canada Year Book 1988[1].

Mackenzie	4241 km	Yukon	3185 km
St. Lawrence	3058 km	Nelson	2575 km
Columbia	2000 km	Saskatchewan	1939 km
Peace	1923 lm	Churchill	1609 km
South Saskatchewan	1392 km	Fraser	1370 km
North Saskatchewan	1287 km	Ottawa	1271 km

6. Make a circle graph to represent the number of years each political party governed the Plains.

Old Party	1905 to 1920	United Workers of the Plains	1921 to 1934
Social Party	1935 to 1970	Progressive Liberal Party	1971 to Present

7. Explain how you would decide to use one type of graph over another type of graph.

For solutions to **Activity 2**, turn to **Appendix A**, **Topic 2**.

[1] Statistics Canada. *Canada Yearbook*. Reprinted with permission of the Minister of Supply and Services Canada.

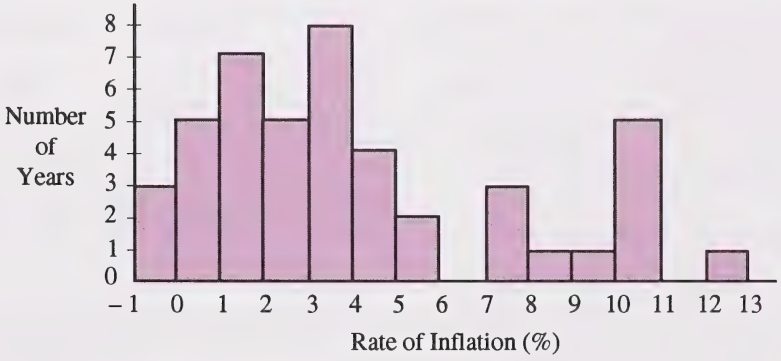
Activity 3

Graph and interpret continuous data using histograms and frequency polygons.

The Histogram

There are only two differences that you will notice between the histogram and the bar graph; the bars in the histogram lie right against each other and the categories are made up of continuous data.

The Number of Years the Inflation Rate was at a Certain Level for Canuckland (1915 to 1987)



On this graph, the continuous data is shown along the bottom. Notice that the numbers run up from −1 to 13; there is no number missing and the numbers are evenly spaced, just like on a ruler. This is referred to as continuous data. All histograms will have their categories written in this form.

All histograms have the following features:

- A title is used to inform the reader what is being shown in the graph.
- Categories are clearly labelled and written using continuous data. The categories must be of equal sizes. They are usually written along the bottom of the graph. This is no longer considered a rule, as it is now possible to find graphs in magazines that do not use the bottom of the graph.
- The bars of the graph must
  - be between the class limits of the continuous data, or must be centred on the midpoint of the continuous data
  - be of equal width
  - be right against each other
- The axes must
  - be labelled with numerical values and identifying label
  - be present on at least one side of the graph
- The scale for the vertical axis must increase by equal values from 0.

A histogram is used when you want a visual comparison of continuous data. The data must be grouped into categories of equal size.

Now look at how to construct a histogram.

The following data is the number of goals scored by players in the Stanley division of the DHL in the 1990/91 season.

50	48	51	21	40	25	17	13	20	9	7	13	10	7	5	7	2	8	0	40	37	43
56	38	23	36	12	20	13	15	11	13	9	5	9	4	8	10	0	3	53	55	32	26
16	19	16	20	16	2	9	9	3	7	7	6	1	1	0	0	40	36	19	30	12	4
7	1	6	15	7	7	6	12	1	7	4	2	2	2	0	44	40	32	15	13	16	7
25	19	18	15	6	8	6	7	5	4	0	1	0									

Step 1: Calculate the range of the data. This is the difference beween the highest and the lowest value in the data.

$$\text{Range} = 56 - 0$$
$$= 56$$

Range is the difference between the highest and the lowest value in the data.

Step 2: Decide on the number of classes. You will want between eight and ten classes.

For this particular case you will use ten classes.

A **class** is the same as an interval.

Step 3: Divide the number of classes into the range as this will give you the class width. If this number is not a whole number, round off to the next highest whole number.

Class width is the number of values placed into each class.

$$\text{Class width} = \frac{\text{Range}}{\text{Classes}}$$
$$= \frac{56}{10}$$
$$= 5.6$$
$$\doteq 6 \quad \text{(rounded to the next highest whole number)}$$

Step 4: Use the class width to decide on the class intervals. Take the lowest value in the data and add the class width to it. This will give you the starting value of the next class. The end of the first class will be one unit less than the value you just calculated. Continue adding on the class width until you have a beginning and an ending value for each class. These values are the class intervals.

Starting point of each class	Ending point of each class
0	5 (6 − 1 = 5)
0 + 6 = 6	5 + 6 = 11
6 + 6 = 12	11 + 6 = 17
12 + 6 = 18	17 + 6 = 23
18 + 6 = 24	23 + 6 = 29
24 + 6 = 30	29 + 6 = 35
30 + 6 = 36	35 + 6 = 41
36 + 6 = 42	41 + 6 = 47
42 + 6 = 48	47 + 6 = 53
48 + 6 = 54	53 + 6 = 59

Check to make sure that all of the values will fit between the class intervals.

A **class interval** is the limit that the data must fit between to belong to a specified class.

Step 5: Make a frequency distribution table using the class intervals and the data that you calculated. This is easier to do if the data is ordered first.

Class	Class Interval	Tally	Frequency
1	0-5	𝍸 𝍸 𝍸 𝍸 𝍸 I	26
2	6-11	𝍸 𝍸 𝍸 𝍸 𝍸 II	27
3	12-17	𝍸 𝍸 𝍸 II	17
4	18-23	𝍸 IIII	9
5	24-29	III	3
6	30-35	III	3
7	36-41	𝍸 III	8
8	42-47	II	2
9	48-53	IIII	4
10	54-59	II	2
		Total	101

Step 6: Selelct an appropriate title for the graph.

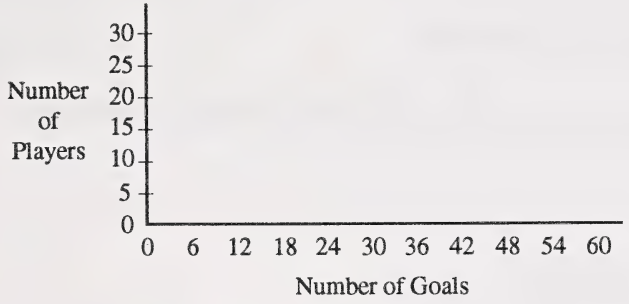
Distribution of the Goal Scorers in the Stanley Division
for the 1990/91 Season

Step 7: Select and label the axes appropriately. Make sure the scale for the vertical axis starts at zero.

The scale should be selected so that the tallest bar will not completely reach the top of the axes. For this particular case, the tallest bar will reach 27; therefore, you will make the scale up to 30.

Step 8: Put the label and lower class intervals, horizontal class limits, along the other axis of the graph.

Distribution of the Goal Scorers in the Stanley Division for the 1990/91 Season



Step 9: Draw in the bars making sure that each bar is the correct length. Also, make sure that all of the bars have the same width, the edge of the bars are pointing directly at the class intervals, and the bars are touching each other.

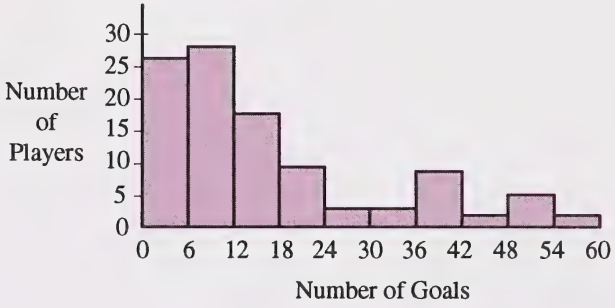
If you were using the midpoints of the classes instead of the class limits, you would have to make sure the bars are centred over the classes instead of pointing at the class limits.

To construct a histogram, do the following:

- Find the range.
- Decide on the number of classes.
- Find the class width.
- Find the class intervals (class limits).
- Make a frequency distribution table.
- Select an appropriate title.
- Select and label an appropriate scale.
- Put in the label on lower class intervals (class limits).
- Draw the bars.

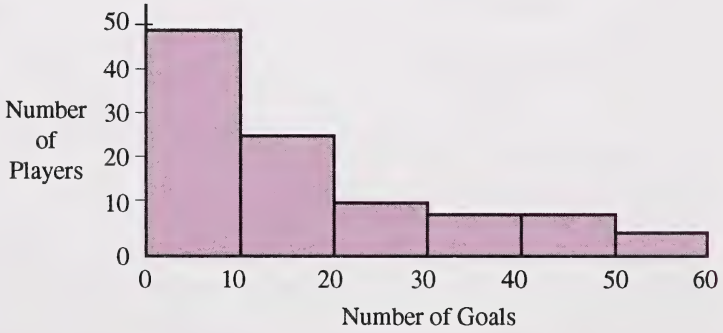
The graph is now complete.

Distribution of the Goal Scorers in the Stanley Division
for the 1990/91 Season



The shape of the histogram is dependent on the number of classes selected. The following is the same graph with different classes and class limits.

Distribution of the Goal Scorers in the Stanley Division
for the 1990/91 Season



Can you believe that these two graphs were derived from the same group of data? Notice how the second graph starts with its high point on the left side and each bar to the right is lower. On the first graph, the second bar is higher than the first bar. In two other places on the graph, as you move from the left to the right, the bars increase in area and height.

The only difference in the construction of these two graphs is in the selection of the number of classes. This is a demonstration of how important it is to be careful with your selection of the number classes.

Do you know how to read information from a histogram? Look at the following example.

Example 1

Find the answers to the following questions from the first histogram.

- How many players scored less than six goals?

  Solution:

  Find the bar or bars that make up the players that have scored less than six goals. In this case it will be the first bar. This bar goes up to 26. Therefore, 26 players scored less than six goals.

- What group of goal scorers is the largest?

  Solution:

  Find the tallest bar. The tallest bar is the second bar. This bar represents players that have scored six to eleven goals.

- Which two groups have the same number of players?

  Solution:

  By examining the graph, you will notice that there are two positions where two bars have the same number of players. The bars that represent 24-29 and 30-35 each have three players. The bars that represent 42-47 and 54-59 each have two players.

- Did more players score less than 12 goals, or 12 or more goals?

  Solution:

  You must decide if the bars to the left of 12 have a greater area or the bars to the right of 12 have a greater area. First, you will have to find the area of each of these two section. Assume that each bar has a width of one; then, you will only need to add the lengths of each bar.

  Left of the 12:

  $$26+27=53$$

  Right of the 12:

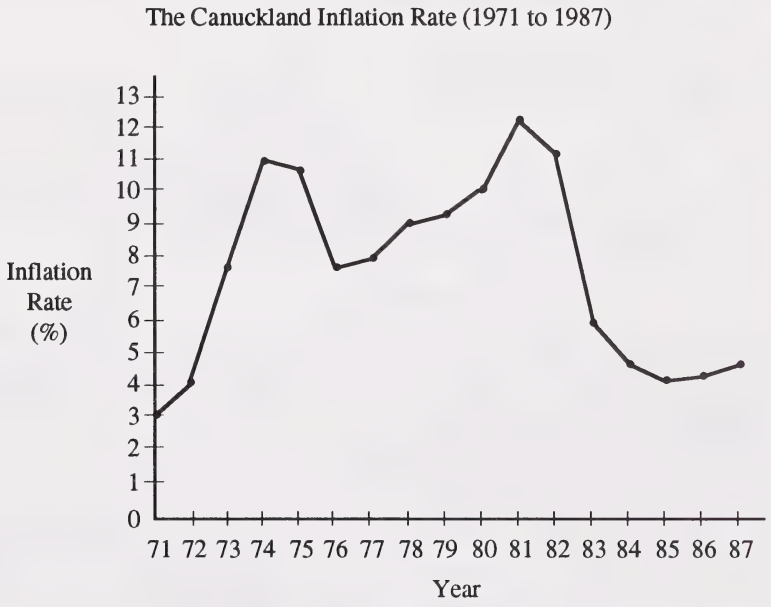
  $$17+9+3+3+8+2+4+2=48$$

  The bars to the left of 12 have a greater area. Thus, more players scored less than twelve goals than those who scored twelve goals or more.

You will not take a look at the frequency polygon.

The Frequency Polygon

The frequency polygon is identical to the histogram except for the bars. The frequency polygon make use of a line connecting points instead of using bars.



Notice that this graph also has a title, axes, labels, and continuous data just like the histogram. The only difference lies on the body of the graph. The bars are replaced with lines that are joining dots. The dots are positions where the centre of the bars would be on the histogram.

All frequency polygons have the following features:

- A title is used to inform the reader what is being shown on the graph.
- The categories must be clearly labelled and written using continuous data. The categories must be of equal sizes. They are usually written along the bottom of the graph.
- The dots of the graph must be between the class limits of the continuous data, or they must be centered on the midpoints of the continuous data.
- The lines of the graph must connect the dots from left to right.
- The axes must
  - be labelled with numerical values and a label
  - be present on at least one side of the graph
- The scale for the vertical axis must increase by equal values from 0.

A frequency polygon is used when you want a visual comparison of continuous data. The data must be grouped into categories of equal size.

When would you use a frequency polygon instead of a histogram? The frequency polygon is usually used when the continuous data is years running in succession. You will usually see a business use this type of graph when they want to compare the difference in interest from one year to the next.

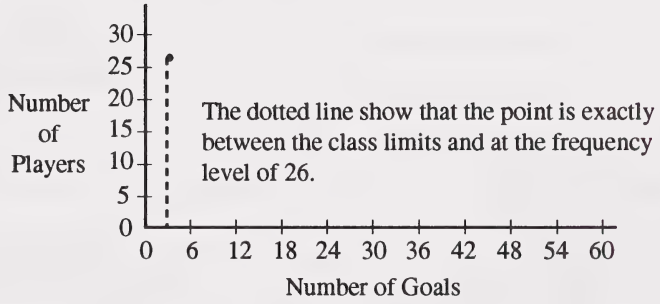
Because so many of the elements of the frequency polygon are identical to the histogram, most of the steps are identical. Only the steps of the frequency polygon that are different from the histogram will be rewritten. The same data will be used.

To construct a frequency polygon, do the following:

- Find the range.
- Decide on the number of classes.
- Find the class width.
- Find the class intervals (class width).
- Make a frequency distribution table.
- Select an appropriate title.
- Select and label an appropriate scale.
- Label the axes appropriately.
- Draw in the points.
- Connect the points.

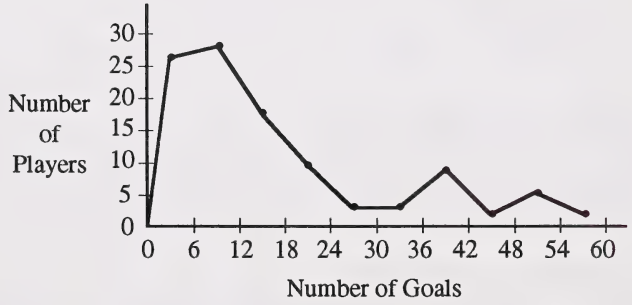
Step 9: Draw points halfway between the class limits and the frequency level. If the classes or the midpoints are used, then the dots will be drawn directly above the classes or midpoints.

Distribution of the Goal Scorers in the Stanley Division
for the 1990/91 Season


Number of Players
30
25
20
15
10
5
0
0 6 12 18 24 30 36 42 48 54 60
Number of Goals
The dotted line show that the point is exactly between the class limits and at the frequency level of 26.

Step 10: Connect the points from the left to right.

Distribution of the Goal Scorers in the Stanley Division
for the 1990/91 Season



The frequency polygon is complete.

Now try some questions on histograms and frequency polygons.

Answer any four of the following questions.

1. Use the following data to construct a histogram.

The Canuckland Price Index

1950	25	1951	27	1952	28
1953	28	1954	28	1955	38
1956	38	1957	29	1958	30
1959	31	1960	31	1961	31
1962	32	1963	32	1964	32
1965	34	1966	35	1967	36
1968	38	1969	39	1970	41
1971	42	1972	44	1973	47
1974	52	1975	58	1976	62
1977	67	1978	73	1979	80
1980	88	1981	100	1982	110
1983	117	1984	122	1985	127
1986	132	1987	138		

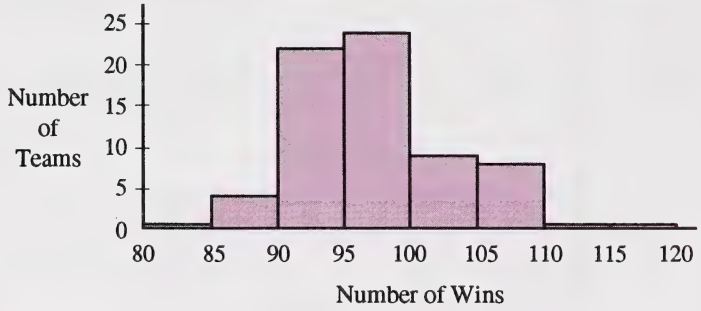
2. Construct a frequency polygon using the following data.

Number of Wins by the Red Deer Titans

1969	52	1970	73	1971	71	1972	70
1973	79	1974	79	1975	75	1976	55
1977	75	1978	76	1979	95	1980	90
1981	60	1982	86	1983	82	1984	78
1985	84	1986	78	1987	91	1988	81

3. Answer the following questions from the histogram.

Number of Wins by the Distance League Pennant Winners
(1911-1978)



a. What is the highest number of possible wins on this graph?

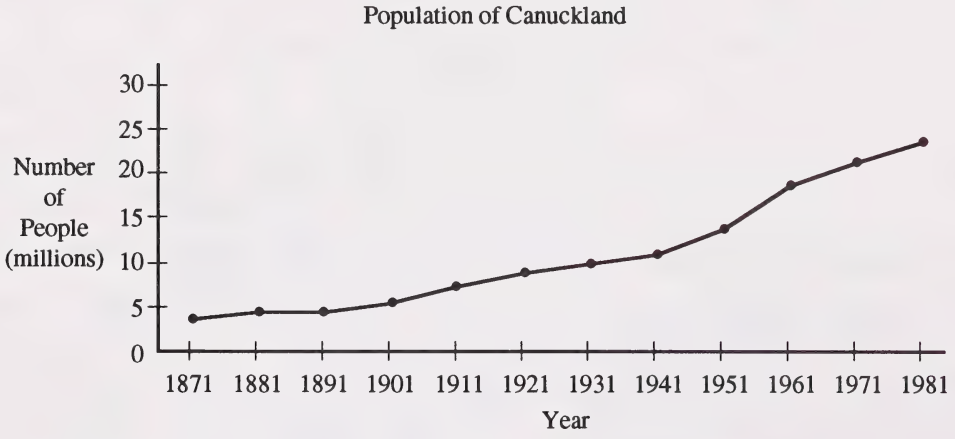
b. Which class has the greatest number of teams?

c. Which class interval has seven teams?

d. How many teams won between 90 and 94 games?

e. How many years or teams have been used in this histogram?

4. Answer the following questions from the following frequency polygon.



a. What was the population in 1911?

b. In what year was the population 10 300 000?

c. In which ten year period was the greatest increase in population?

d. How may this data have been collected?

e. In which ten year period did the population reach over twenty million?

5. Explain why two histograms based on the same data may look completely different.

For solutions to **Activity 3**, turn to **Appendix A, Topic 2**.

If you require help, do the Extra Help section.

If you want more challenging explorations, do the Extensions section.

} You may decide to do both.

Extra Help

More About Tabulating Data

Joe made a survey of the sports that students played most at Griffiths High School. The data collected from the randomly selected sample is displayed as follows.

S	(S)	(B)	F	S	T
H	B	S	(S)	(B)	B
B	T	H	B	H	H
F	T	H	B	F	F
F	B	H	T	F	F
(S)	(B)	F	T	F	B

S – swimming
(S) – skiing
B – basketball
T – tennis
(B) – baseball
H – hockey
F – football

Patterns and trends can be more easily seen if the data is displayed in an organized manner.

A simple way to organize data is to use a frequency table.

A frequency table uses three headings to record data. Check the table by adding the frequencies. The total should be 36.

Frequency Table

Sport	Tally	Frequency
Swimming	III	3
Skiing	III	3
Basketball	𝍸 II	7
Tennis	𝍸	5
Baseball	III	3
Hockey	𝍸 I	6
Football	𝍸 IIII	9

From this table, you can see at a glance, that football is more popular than any other sport. You can see that swimming, skiing, and baseball are the least popular sports. You can also calculate that $\frac{9}{36} \times 100 = 25\%$ students preferred football.

Try answering these questions.

1. Place the following data in a frequency table.

Ages of Thirty Students Randomly Selected at Griffiths High School

15	15	14	17	16	14
18	16	17	18	18	16
19	18	16	17	16	14
16	17	16	14	16	15
17	15	17	16	15	15

2. Using the table in question 1, answer these questions.

   a. How many students surveyed are sixteen years old?

   b. How many more sixteen-year-old students were surveyed than fifteen-year-old students?

   c. What percent of the students surveyed are sixteen years old?

For solutions to **Extra Help** turn to **Appendix A, Topic 2.**

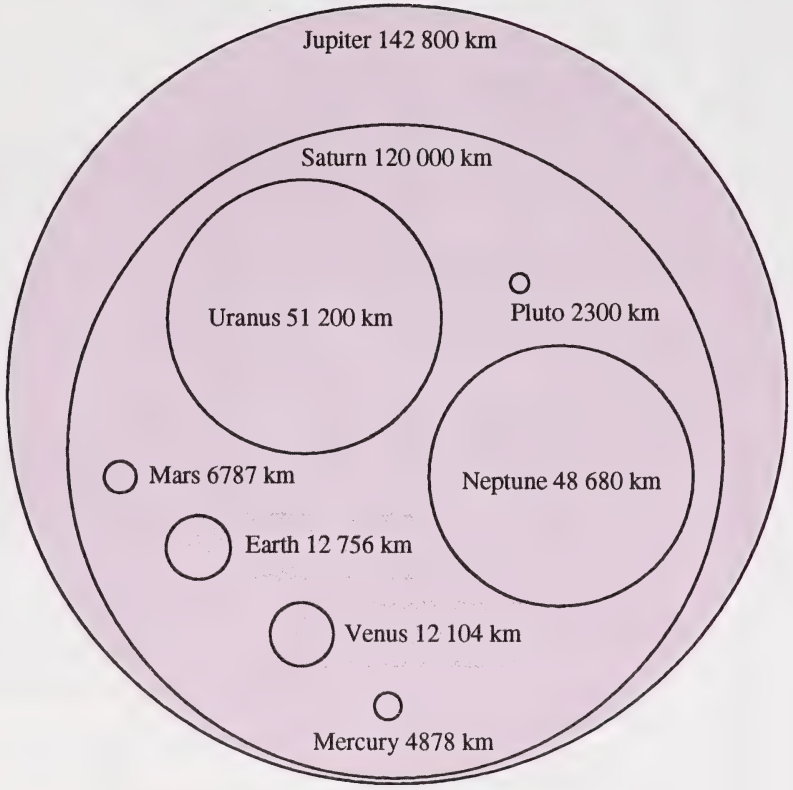
Extensions

More on Pictographs

Did you realize that there is another type of pictograph?

You may have seen this type of pictograph before but did not realize what is was. Take a closer look at the following **single picture pictograph.**

The Diameters of the Planets



Source: 1992 Canadian Global Almanac

Scale: $1\text{ cm} = 1.2 \times 10^4$ km

There are two things that are noticeable about this type of graph.

- There is only one picture for each category.
- The size of the picture is different for each category.

The key for making the picture lies in the scale.

For this case, the scale 1 cm equals 1.2×10^4 km means that 1 cm on the pictograph will represent 1.2×10^4 km on the actual planet. If you measure the diameter of one of the circles and multiply that number by 1.2×10^4 km, you get an answer that represents the actual size of the planet.

This type of picture will have the following features.

- A title to inform the reader what is being shown in the graph.
- The items or categories must be clearly defined. These will be written somewhere near or on the picture.
- A scale must be present to tell the reader what each picture represents.
- All pictures must be in correct proportion to each other as denoted by the scale and the actual amount of each item.
- The actual amount of each item must be clearly written with the category label.

This type of pictograph is used when you want to make a visual comparison of similar data. The data is grouped into separate distinct categories.

Look at the symbol for Neptune. Its diameter is 4 cm. The actual diameter of Neptune would be about

$$4\times1.2\times10^4 \doteq 4.8\times10^4$$
$$\doteq 48\,000 \text{ km}$$

Now look at how to construct this type of pictograph. You will use the same data that was used to make the pictograph earlier in the topic.

Actual Number of CDs Produced by Arnie's Sound

Artist	Size of Population
Belinda Aisle	300 000
Brain Odums	150 000
Dadonna	400 000
Flatwood Joe	250 000
Men Out of Work	125 000
Neil Zirconia	175 000
U3$\frac{1}{2}$	750 000
Von Jodi	350 000
Total	2 500 000

Step 1: Select an appropriate title for the pictograph.

You may use the same title as the one used to name the data table.

Step 2: Select the symbol to be used as a picture in the graph.

Since compact discs are circular objects, use circles.

Step 3: Decide on the scale, and write this on the graph.

This is done by looking at how much space you have to work with and how large the largest item is. Divide the largest item into the amount of space you have as this will give you the scale. In this case, use a scale of 1 mm = 12 000 CD's.

Place the following title on the graph.

Actual Number of CDs Produced by Arnie's Sound.

Step 4: Using the scale, find the measure of the diameter of each circle. Set up a proportion to find these measures. Round to the nearest whole number.

Belinda Aisle

$$\frac{1\text{ mm}}{12\,000\text{ CDs}} = \frac{n}{300\,000\text{ CDs}}$$

$$n = \frac{1\text{ mm} \times 300\,000\text{ CDs}}{12\,000\text{ CDs}}$$

$$= 25\text{ mm}$$

Brian Odums

$$\frac{1\text{ mm}}{12\,000\text{ CDs}} = \frac{n}{150\,000\text{ CDs}}$$

$$n = \frac{1\text{ mm} \times 150\,000\text{ CDs}}{12\,000\text{ CDs}}$$

$$= 13\text{ mm}$$

Dadonna

$$\frac{1\text{ mm}}{12\,000\text{ CDs}} = \frac{n}{400\,000\text{ CDs}}$$

$$n = \frac{1\text{ mm} \times 400\,000\text{ CDs}}{12\,000\text{ CDs}}$$

$$= 33\text{ mm}$$

Flatwood Joe

$$\frac{1\text{ mm}}{12\,000\text{ CDs}} = \frac{n}{250\,000\text{ CDs}}$$

$$n = \frac{1\text{ mm} \times 250\,000\text{ CDs}}{12\,000\text{ CDs}}$$

$$= 21\text{ mm}$$

Men Out of Work

$$\frac{1\text{ mm}}{12\,000\text{ CDs}} = \frac{n}{125\,000\text{ CDs}}$$

$$n = \frac{1\text{ mm} \times 125\,000\text{ CDs}}{12\,000\text{ CDs}}$$

$$= 10\text{ mm}$$

Neil Zirconia

$$\frac{1\text{ mm}}{12\,000\text{ CDs}} = \frac{n}{175\,000\text{ CDs}}$$

$$n = \frac{1\text{ mm} \times 175\,000\text{ CDs}}{12\,000\text{ CDs}}$$

$$= 15\text{ mm}$$

U3$\frac{1}{2}$

$$\frac{1\text{ mm}}{12\,000\text{ CDs}} = \frac{n}{750\,000\text{ CDs}}$$

$$n = \frac{1\text{ mm} \times 750\,000\text{ CDs}}{12\,000\text{ CDs}}$$

$$= 63\text{ mm}$$

Von Jodi

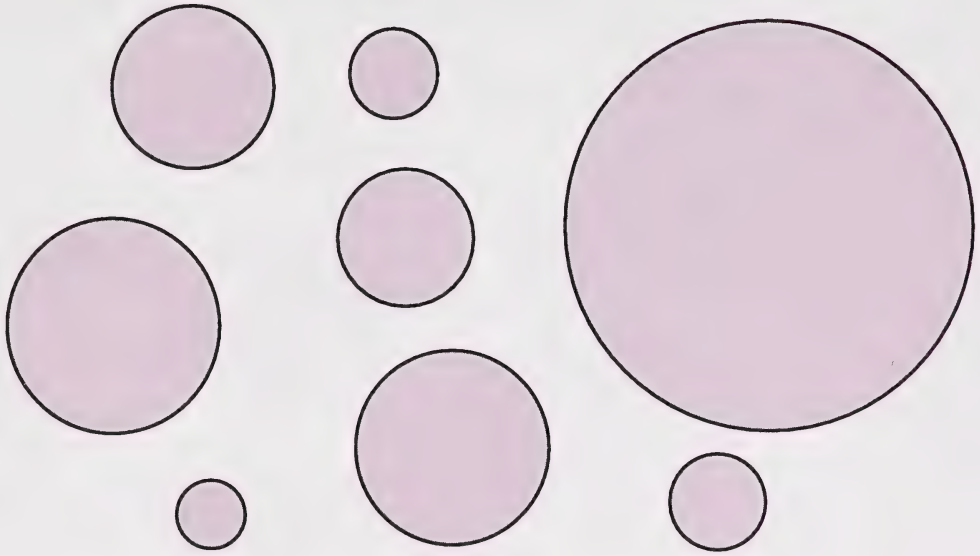
$$\frac{1\text{ mm}}{12\,000\text{ CDs}} = \frac{n}{350\,000\text{ CDs}}$$

$$n = \frac{1\text{ mm} \times 350\,000\text{ CDs}}{12\,000\text{ CDs}}$$

$$= 29\text{ mm}$$

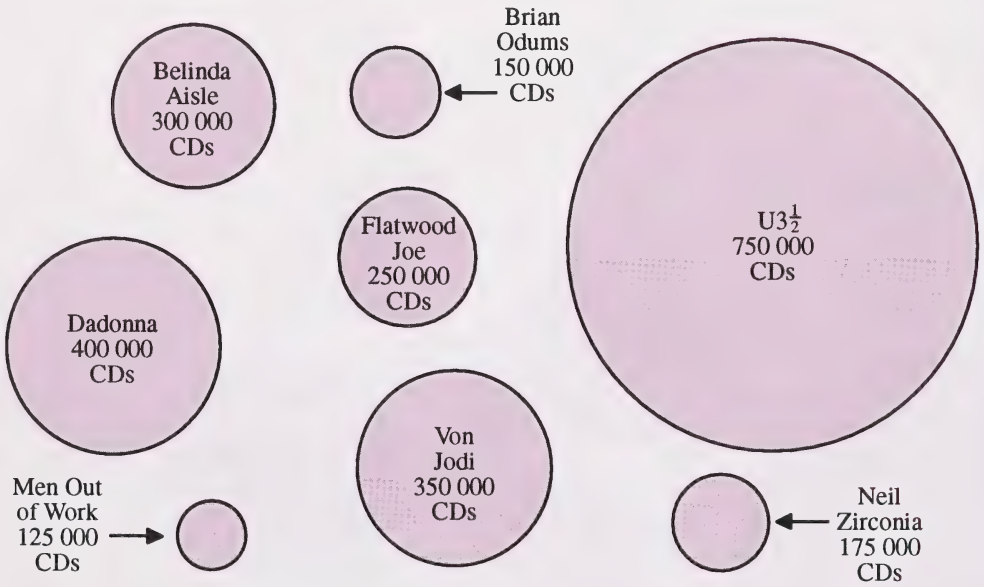
Step 5: Draw the pictures on the graph.

Actual Number of CDs Produced by Arnie's Sound



Step 6: Put the correct labels and amount on each symbol or beside each symbol.

Actual Number of Discs Produced by Arnie Sounds
Scale: 1mm = 12 000 CD's



The graph is now complete.

The symbols in this graph differ from the first graph. On this graph, the symbols are kept as separate items. On the first graph, the symbols were placed on top of each other.

Symbols will only be placed on top of each other when one or two of the items are so much larger than the other items that it becomes necessary to make them as large as possible so that the small items will have a recognizable size of the graph.

How to construct a pictograph.

- Select an appropriate title.
- Select the picture or symbol.
- Find the scale.
- Find the measure of each item.
- Draw the pictures or symbols.
- Label the pictures or symbols.

Now try the following questions.

Answer any five of the following questions.

Answer questions 1 to 4 using the graph that was just made.

1. Which item has the largest picture or symbol?

2. What is the amount on the smallest picture or symbol?

3. What is the scale for the graph?

4. How many times larger is the U3$\frac{1}{2}$ circle than the Neil Zirconia circle? (Use diamter or radius.)

5. Make a single picture or single symbol pictograph using the following data.

The Amount of Energy in Food

Food	Amount	Calories
Cheddar cheese	45 g	182
Hard vanilla ice cream	250 mL	143
Mayonnaise	15 mL	99
Chocolate bars	30 g	135
Potato chips	10 chips	114
Pop	200 mL	78
Chocolate coated peanuts	30 g	157

6. Make a single picture or single symbol pictograph using the following data.

Canada's Largest Waterfalls

Waterfall	Vertical Drop
Takakkaw Falls	503 m
Della Falls	440 m
Hunlen Falls	253 m
Panther Falls	183 m
Helmcken Falls	137 m
Bridal Veil Falls	122 m
Virginia Falls	90 m

Source: Canada Yearbook, 1988, Statistics Canada[1]

For solutions to **Extensions**, turn to **Appendix A, Topic 2**.

[1] Statistics Canada. *Canada Yearbook*. Reprinted with permission of the Minister of Supply and Services Canada.

Topic 3 Analysis of Data: Mean, Median, Mode

Introduction

The mean, median, and mode are the big three statistics. If you see a statistic recorded, it will likely be one of these three. They are all called averages, but each measures something very different. It is important to understand each and the difference between them.

What Lies Ahead

Throughout the topic you will learn to

1. define the measures of central tendency, and determine the most suitable measure of central tendency for a given set of data
2. determine the mean, median, and mode for discrete raw data
3. determine the mean, median, and mode for tabular data
4. determine the range for a given set of data

Now that you know what to expect, turn the page to begin your study of analysis of data: mean, median, mode.

Exploring Topic 3

Activity 1

Define the measures of central tendency, and determine the most suitable measure of central tendency for a given set of data.

In this topic you are going to be looking at three measures of central tendency: the **mean**, the **median**, and the **mode**. Take a look at what each of these are.

The **mean**, or arithmetic mean, is more commonly known as the average. The mean is found by adding all the values of the data and dividing this sum by the number of elements of data.

One or two extreme values can make the value of the mean misleading.

The **median** is the middle value from the rank-ordered data. The data first must be arranged in ascending order. Then the middle value is selected. If there is an even number of elements, there will not be a middle value. In this case, you take the average of the two middle elements.

Extreme values in the data will not affect the median.

The **mode** is the value that appears most frequently in the data. The mode is most easily found when the data is organized in ascending order.

It is possible for data to have more than one mode or no mode. If a set of data has two modes, it is called bimodal.

Like the median, the mode is not influenced by extreme values in the data since only a small number of values is used to find the mode.

The following are important concepts to remember:

- **Ascending order** means to arrange numbers so that they are written from the smallest number to the largest number.
- mean = $\frac{\text{sum of all elements of data}}{\text{number of elements of data}}$
- median = the middle value from organized data
- mode = the value which occurs most often within the data

The following guidelines may help you decide which measure of central tendency is most suitable for any given situation.

- The mean is most appropriate when the data has no extreme values.
- The median is most appropriate when there are a few extreme values. The median is least affected by extreme values.
- The mode is most appropriate when you are looking for the value or values that occur most frequently.

The mean is the best indicator when you are finding

- a student's grade average
- average gas consumption
- average annual rainfall or snowfall
- a hockey goalkeeper's average of goals allowed per game
- points per game scored by a basketball player
- a baseball player's hitting average
- bowling averages

The median is the best indicator when you are finding

- wages for a company
- competition times for athletes
- average test marks for a class
- number of defective items being produced
- attendance records

The mode is the best indicator when you are finding

- hat sizes
- dress sizes
- shoe sizes

Example 1

Use either mean, median, or mode to indicate the best meaning of **average** in each statement given.

- The average seventeen-year-old student is in grade eleven.

  Solution:

  In this statement average refers to the mode since most seventeen-year-olds are in grade eleven.

- The average rollerskate size rented at Rollerland is size eight.

  Solution:

  In this statement average refers to the mode since the most frequently rented size is size eight.

- An average of three fish per day was caught at the lake.

  Solution:

  In this statement average refers to the mean, providing there are no extreme daily catches, like seventeen fish caught one day and zero on several other days.

- The company's average annual salary is $17 000.

  Solution:

  In this statement average refers to the median, since several extreme values may exist.

- Karen's overall report card average was 69%.

  Solution:

  In this statement average refers to the mean, since there should not be any extreme values.

Now try the following questions.

Answer question 1 and any two of questions 2, 3, and 4.

1. Which measure of central tendency (mean, median, or mode) would best describe each of the following data?

   a. an average weekly salary
   b. a curling average
   c. the amount of rainfall a town receives
   d. the average mark on a test in a class
   e. the cost of gasoline in a town
   f. the most popular dress size
   g. the number of crackers in a box
   h. the number of goals scored per game in soccer
   i. the number of hours spent studying each night
   j. the number of hours spent watching television each night
   k. the average number of people per family

2. Is it necessary to put data into ascending order before finding the mean?

3. Is it necessary to organize data before finding the mode?

4. What is the first thing you should do when you are finding the median for a group of data?

For solutions to **Activity 1**, turn to **Appendix A**, **Topic 3**.

Activity 2

Determine the mean, median, and mode for discrete raw data.

How do you calculate the measures of central tendency?

The mean, or **arithmetic mean**, is calculated by dividing the sum of the data by the number of values present in the data.

Take a look at this case. The number of defective compact discs that were found each day from a random selection manufactured at Arnie's Sound was as follows.

6	7	8	8	8	9	11	11
13	14	15	15	16	17	17	17

The mean is as follows:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots x_n}{n}, \text{ where } n \text{ is a natural number}$$

$$= \frac{6+7+8+8+8+9+11+11+13+14+15+15+16+17+17+17}{16}$$

$$= \frac{192}{16}$$

$$= 12$$

Arnie's Sound found a mean of twelve defective CDs per day.

Note that

- x_1 means the first data element
- x_2 means the second data element
- x_3 means the third data element
- x_n means the last data element
- $\bar{x}$ is the mathematical symbol for mean

The **median** is the middle value or the average of the two middle values after the data has been arranged in ascending order.

Count the number of values in the data. If there is an **odd** number of values, the median will be the middle value. If there is an **even** number of values, the median will be the average of the two middle values.

Take another look at the Arnie's Sound case.

This data has already been arranged in ascending order so you will only need to find the two middle values. This group of data has an even number of values, so the median will be the average of the two middle values.

first value ↓

6	7	8	8	8	9	11	(11)
(13)	14	15	15	16	17	17	17

↑ last value

The two middle values are 11 and 13. The average of the these two values is as follows:

$$\frac{11+13}{2} = \frac{24}{2}$$

$$= 12$$

The median value of defective CDs is 12.

If you are still having trouble with the median, turn to the **Extra Help** section at the end of this topic.

The **mode** is the value or values that occur most frequently.

The mode is most easily found when the data is organized in ascending order. The data for Arnie's Sound has already been organized. Count the number of times each value appears in the data.

(6) (7) (8 8 8) (9) (11 11)

three times

(13) (14) (15 15) (16) (17 17 17)

three times

The value 8 appears three times. The value 17 also appears three times. Therefore, this group of data is bimodal and the modes are 8 and 17.

Bimodal means that the sample has two modes just as a has two wheels and just as a has two wings on each side.

What will happen to the measures of central tendency if one of the values is changed?

Change the 16 to 64 and see how this affects the three measures.

First see how the changed value affects the mean.

$$\bar{x} = \frac{6+7+8+8+8+9+11+11+13+14+15+15+64+17+17+17}{16}$$

$$= \frac{240}{16}$$

$$= 15$$

The value of the mean has increased from 12 to 15.

Now see how the changed value affects the median.

First the data will have to be rearranged in ascending order.

6	7	8	8	8	9	11	(11)
(13)	14	15	15	17	17	17	64

Now locate the two middle terms. They are still 11 and 13. Therefore, the median remains at 12.

Now, look through the data and find the mode.

There are still the two modes of 8 and 17, since both appear three times in the sample.

What will happen if another value is added to the data?

- Take the original data and add the value 29 to it. What is the mean, the median, and the mode?

  The solution is as follows:

  $$\bar{x} = \frac{221}{17}$$

  $$= 13$$

  median = 13
  mode = 8 and 17

- Take the original data and add the value 12 to it. What is the mean, the median, and the mode?

  The solution is as follows:

  $$\bar{x} = \frac{204}{17}$$

  $$= 12$$

  median = 12
  mode = 8 and 17

- Take the original data and add the value 20 to it. What is the mean, the median, and the mode?

  The solution is as follows:

  $$\bar{x} = \frac{212}{17}$$

  $$= 12$$

  median = 13
  mode = 8 and 17

Summarize what you have discovered from doing these three cases.

Did you come to the same conclusions as the following?

The mean will change if the value that is added to the data is not the same as the mean.

The median will usually change if the value that is added to the data is not the same as the median.

The mode will change only if the value added to the data is the same as one of the modes or the same as one of the values that occurred one less time than the mode.

If you have access to a videocassette recorder, you may view the *Math Wise* video titled **Predicting – Averages**. This program discusses the influence of extreme values on the average.

Try any four of the following questions.

1. Find the mean, the median, and the mode for the following data. Give your answer rounded to one decimal place if necessary.

   35 34 39 35 30 34 32 38 34 39

2. The employees for Arnie's Sound have monthly wages of $2100, $2400, $1700, $2100, $2100, $3200, $2500, $2500, and $5200. Which type of measure will give the best average for this data? What is the average?

3. Using the following data, which measure will give the best average of the length of one day in the solar system? What is the average?

Planet	Length of Day (hours)
Mercury	1416
Venus	5834
Earth	24
Mars	26
Jupiter	9
Saturn	10
Uranus	16
Neptune	19
Pluto	154

4. Each of the twenty-one students in a science class measured the mass of the same object using an equal arm balance. The results are shown as follows.

236 g	238 g	242 g	239 g	238 g	237 g	239 g
240 g	236 g	240 g	238 g	235 g	254 g	239 g
237 g	238 g	237 g	236 g	242 g	241 g	243 g

What type of measure best describes the average of this data? What is the average when rounded to the nearest tenth?

5. The G&E Department store has twenty-one hats. The sizes are given as follows. Which measure gives the best average of the sizes? What is the average?

180 mm	165 mm	185 mm	185 mm	190 mm
190 mm	175 mm	185 mm	180 mm	185 mm
170 mm	185 mm	195 mm	195 mm	185 mm
195 mm	185 mm	180 mm	185 mm	175 mm
200 mm				

For solutions to **Activity 2**, turn to **Appendix A**, **Topic 3**.

Activity 3

Determine the mean, median, and mode for tabular data.

How do you find the measures of central tendency from tables, stem-and-leaf plots, and box plots?

Start by looking at the stem-and-leaf plot.

Recall from the last topic that all the values of the data are given and put in ascending order in the stem-and-leaf plot. This saves time when you are looking for the median and the mode.

Take a look at a particular case.

Final Marks for Mathematics 13, January 1991

3	4 5 5
4	3 7
5	0 1 2 6 6 9
6	2 2 5 5 7 7 7 8 8 9
7	2 5 5 8 8 8
8	1 6 9
9	2

This stem-and-leaf plot gives the final marks for a Mathematics 13 class. All the marks are given in the plot.

The numbers on the left will be the digit in the tens place, and the numbers on the right will be the digits in the ones place. For example, the numbers in the first row are the values 34, 35, and 35. The 3 is placed in front of all the digits given on the right side of the plot.

The mean is the sum of all the values in the stem-and-leaf plot divided by the number of elements.

In this case the number of elements is 31.

$$\bar{x} = \frac{34 + 35 + 35 + 43 + \ldots + 92}{31}$$

$$\bar{x} = \frac{1982}{31}$$

$$\bar{x} \doteq 63.935\,483\,87$$

$$\bar{x} \doteq 63.9$$

The mean is 63.9%.

The median is the middle value of the data after it is arranged in ascending order.

Since the data is already in ascending order, you only need to find the middle value. In this particular case, there are 31 values. Therefore, the sixteenth value, counted from the beginning or the end, will be the median.

Final Marks for Mathematics 13, January 1991

3	4 5 5
4	3 7
5	0 1 2 6 6 9
6	2 2 5 5 7 7 7 8 8 9
7	2 5 5 8 8 8
8	1 6 9
9	2

The sixteenth value.

The median is 67%.

The mode is the value that occurs most frequently.

To find the mode, count the number of times that different ones digits occur in each row. Whichever digit occurs most frequently in any row will be the mode.

By examining the plot you will notice that there are three 7s in row 4 and three 8s in row 5.

Final Marks for Mathematics 13, January 1991

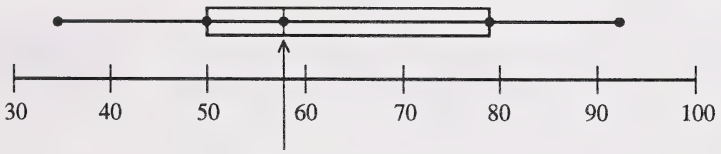
3	4 5 5
4	3 7
5	0 1 2 6 6 9
6	2 2 5 5 7 7 7 8 8 9
7	2 5 5 8 8 8
8	1 6 9
9	2

Since 67 and 78 occur three times each, the modes are 67 and 78.

This data is bimodal since there are two modes.

A box plot gives very little information. Since the centre line of the box plot will pass through the median, the median can be estimated from the box plot; but the box plot does not give enough information to calculate the mean and the mode.

The following box is constructed from the data given in the stem-and-leaf plot. What is the median?



This line represents the median.

The median is about 57%.

Try the following questions.

1. The followng stem-and-leaf plot gives the top 30 scores for a Wolf Greens Amateur golf tournament. Find the three measures of central tendency for the data.

The Top 30 at Wolf Greens

27	6 8
28	0 0 3 6 6 9
29	0 1 1 1 5 5 7 9 9
30	2 2 5 5 8 8 8
31	0 0 6 7 7 9

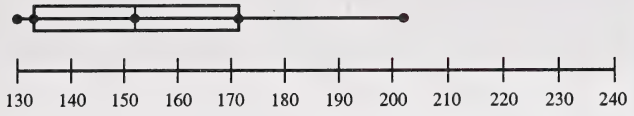
2. Find the three measures of central tendency for the following stem-and-leaf plot.

The Points Earned by the Banff Boomers, 1955 to 1990

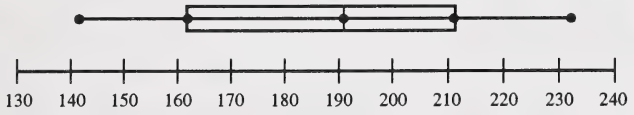
7	5 5 9
8	1 2 3 5 7 7
9	0 1 2 2 2 2 3 4 6 7 8 8 9
10	0 3 3 3 4 7 9
11	3 5 8
12	0 7 9
13	2

3. Find the median for the following box plots.

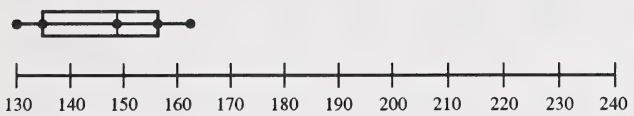
a.



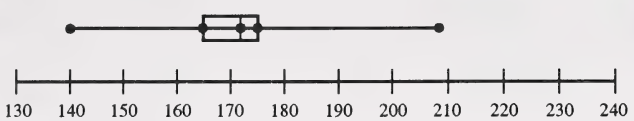
b.



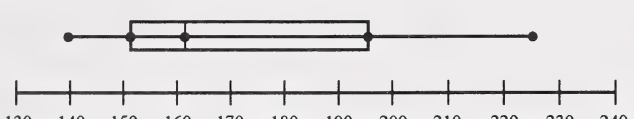
c.



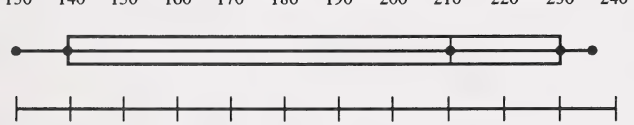
d.



e.



f.



For solutions to **Activity 3**, turn to **Appendix A**, **Topic 3**.

Activity 4

Determine the range for a given set of data.

Up to this point, you have been working with the measures of central tendency. Now you are going to work with a measure of dispersion called the range.

The **range** is the difference between the highest and the smallest value found in a group of data.

The range is easily found using the following procedure.

- Find the largest value in the data.
- Find the smallest value in the data.
- Subtract the smaller value from the larger value.

For example, the number of defective compact discs manufactured at Arnie's Sound that were found from a random selection each day was as follows.

6	7	8	8	8	9	11	11
13	14	15	15	16	17	17	17

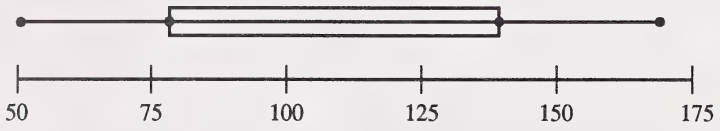
The largest value is 17 and the smallest value is 6.

$$\text{Range} = 17 - 6$$
$$= 11$$

The range for this case is 11 CDs.

How do you find the range from a box-and-whisker plot?

Look at the following box-and-whisker plot.



Here, you look for the location of the ends of the whiskers. The upper end is at about 170 and the lower end is at about 50.

$$\text{Range} = 170 - 50$$
$$= 120$$

The range for the given box-and-whisker plot is 120.

Try the following questions.

1. Find the range for each of the following groups of data.

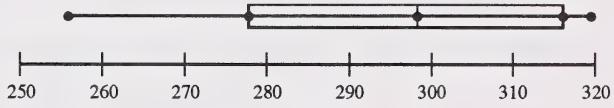
   a. Points Scored in the 1988/1989 DBA Season

2868	1938	1669	2397	1932	1621	1704
2275	2000	1607	2264	1641	1653	1674
1574	1446	1446	1805	1938	2185	2268

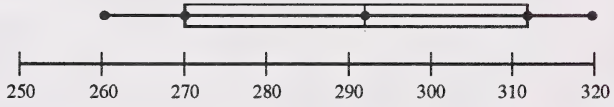
   b. Goals Scored by the Scoring Champions DHL

2	3 6 6 8 9
3	0 1 3 3 3 5 6 6 8 8 9 9
4	0 1 3 4 6 7 7 7 9 9
5	0 0 2 3 4 5 5 6 6
6	0 2 6 8
7	0 1 3 6
8	7
9	2

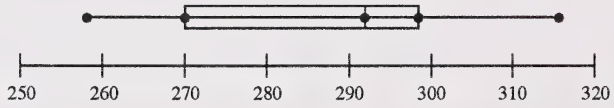
2. a. Wolf Greens Amateur Championship Scores in 1987



   b. Wolf Greens Amateur Championship Scores in 1988



   c. Wolf Greens Amateur Championship Scores in 1989



For solutions to **Activity 4**, turn to **Appendix A**, **Topic 3**.

If you require help, do the Extra Help section.

If you want more challenging explorations, do the Extensions section.

} You may decide to do both.

Extra Help

Finding the Median

The same procedure is always used for finding the mean and the mode, but there are two methods for determining the median.

Depending on the situation, you must select the correct method and then apply it.

It is easy to decide which method to use. To start, count the number of values in the data. If there is an odd number of values, then you will use the first method. If you have an even number of values, then you will use the second method.

Method 1

You will use this method when there is an odd number of values.

Step 1: Count the number of values.

Step 2: Arrange the values in ascending order.

Step 3: Divide the number of values by two and round the number up to the next whole number.

Step 4: Using the number that you found in Step 3, count off that many values from the start of the arranged data.

Step 5: Using the number that you found in Step 3, count off that many values from the end of the arranged data.

Step 4 and Step 5 should give you the same value. If not, go back to Step 1 and go through the steps again.

If Step 4 and Step 5 do give you the exact same value, then that value will be the median.

Method 2

You will use this method when there is an even number of values.

Step 1: Count the number of values.

Step 2: Arrange the values in ascending order.

Step 3: Divide the number of values by two.

Step 4: Using the number that you found in Step 3, count off that many values from the start of the arranged data. You will use this value and the next value.

Step 5: Using the number that you found in Step 3, count off that many values from the end of the arranged data. Take this and the previous value.

Step 4 and Step 5 should give you the exact same values. If not, go back to Step 1 and go through the steps again.

If Step 4 and Step 5 do give the same values, then these values will be used in Step 6.

Step 6: Take the average of the two values.

The solution to the calculation in Step 6 will be the median for the data.

Try the following questions.

1. Which methods would you use to find the median if the data has the following number of values?

   a. 23 b. 105

   c. 56 d. 186

2. Using Step 3, what number results for each group of data that follows? (Note: First you must decide which method you need to use.)

   a. 23, 65, 39, 92, 28, 63, 27, 83, 65, 73

   b. 184, 458, 382, 826, 382, 572, 865, 184, 34

   c. 12, 382, 7783, 239, 22, 264, 2002, 49, 340, 67, 49

3. What is the median for each of the following groups of data?

   a. 8, 6, 4, 7, 3, 9, 6, 3, 34, 19, 4, 15, 8

   b. 12, 27, 24, 18, 20, 22, 13, 16, 15, 23, 45, 1

   c. 102, 116, 89, 115, 121, 98, 92, 107, 113, 114, 97

For solutions to **Extra Help**, turn to **Appendix A**, **Topic 3**.

Extensions

Another Type of Dispersion – Mean Deviation

Sometimes it is not important to know on which side of the mean a particular value lies, but how far that value is from the mean.

The measurement of dispersion that gives you this information is called the **mean deviation**.

The mean deviation is found by dividing the number of values into the sum of the absolute values of the differences between the values and the mean.

See how this is done.

You usually do this type of calculation in tabular form as shown.

Value	Mean	Difference	Deviation
6	10	-4	$\lvert-4\rvert = 4$
6	10	-4	$\lvert-4\rvert = 4$
7	10	-3	$\lvert-3\rvert = 3$
8	10	-2	$\lvert-2\rvert = 2$
10	10	0	$\lvert0\rvert = 0$
13	10	3	$\lvert3\rvert = 3$
14	10	4	$\lvert4\rvert = 4$
16	10	6	$\lvert6\rvert = 6$

$$\overline{M} = \frac{26}{8}$$
$$= 3.25$$

The mean deviation of this group of data is 3.25.

When you have $\lvert-3\rvert$, this means the absolute value of -3.

$$\lvert-3\rvert = 3$$

When you have $\lvert+3\rvert$ or $\lvert3\rvert$, this means the absolute value of positive 3 or just 3.

$$\lvert+3\rvert = 3 \text{ or } \lvert3\rvert = 3$$

The difference is each value minus the mean. The deviation is the absolute value of the difference. To find the mean deviation or $\overline{M}$, use $\overline{M} = \frac{\lvert\text{difference}\rvert}{\text{number of values}}$. $\overline{M}$ is the symbol for mean deviation.

From the previous example you can see that the steps for finding the mean deviation are as follows.

Step 1: Write the values of the data down in the first column of the table.

Step 2: Find the mean for the data and write that number down in the second column opposite each value of the table.

Step 3: Subtract the number in the second column from the number in the first column, and write this difference in the third column.

Step 4: Take the absolute value of the number in the third column, and write this number in the fourth column.

Step 5: Find the sum of all the numbers in the fourth column and write this as the total at the bottom of the column.

Step 6: Count the number of values and divide this number into the total.

The number that you just found is the mean deviation for that group of data.

The general formula for mean deviation is written as $\overline{M} = \frac{\sum |x - \bar{x}|}{n}$,

where $\overline{M}$ = the mean deviation,
$\sum$ = the sum of,
x = the value,
$\bar{x}$ = the mean, and
n = the number of values.

The mean deviation indicates how spread out or dispersed the values are from the mean. The greater the mean deviation, the greater the dispersion.

Try the following questions that deal with mean deviation.

Answer any two of the following questions. Round the means to the nearest whole number.

1. Find the mean deviation for each of the following companies.

Number of Vehicles Produced per Month in 1989 (in 100 000s)

	Tonkyo	Nosson	Sozokki
January	17	13	3
February	17	23	20
March	17	15	19
April	17	21	22
May	17	14	21
June	17	16	17

How does the mean deviation compare to the range for these three companies?

2. a. Find the mean deviation for each of the following teams. Express all means or $\bar{x}$ rounded to the nearest whole number.

Number of Hits by the Top Six Players in 1989

Red Deer	St. Paul	Jasper
184	164	179
164	160	178
152	155	163
116	133	163
114	128	144
100	125	143

b. Which team has the smallest mean deviation?

3. a. Find the mean deviation for each of the following temperatures.

Average Temperature for Selected Canadian Cities

	Red Deer	Quebec	Vancouver
January	−12	−10	3
February	−7	−9	5
March	−4	−3	6
April	3	6	8
May	9	13	12
June	14	18	14

b. Which city has the largest mean deviation?

For solutions to **Extensions**, turn to **Appendix A**, **Topic 3**.

Topic 4 Drawing Inferences

Introduction

You have studied several methods for interpreting data. This resulted in some patterns which were not obvious from the information collected originally.

In this section you will see how these patterns are used in order to make predictions that you feel confident about.

What Lies Ahead

Throughout the topic you will learn to

1. compare and contrast two or more sets of data that address similar situations
2. identify the vocabulary of probability as it applies to the likelihood of the occurrence of an event within a sample
3. make a probability statement and show how it relates to a sample and the population
4. comment on the confidence of predictions that are based on the analysis of a sample

Now that you know what to expect, turn the page to begin your study of drawing inferences.

Exploring Topic 4

Activity 1

Compare and contrast two or more sets of data that address similar situations.

Occasionally, it is interesting to compare two separate groups of data in the same diagram. One method of doing this is to use **back-to-back** stem-and-leaf plots.

The following diagram compares the top twenty hitters from the Distance Baseball League with the top twenty hitters from the Canuckland Baseball League.

Distance and Canuckland League Home-Run Totals for 1991

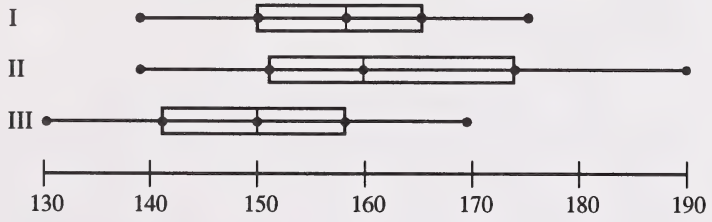
Top 20 DL Home-Run Totals		Top 20 CL Home-Run Totals
4 5 6 6 6 6 7 7 7 7 7 8 9	2	
5 5 5 5 6 7 9	3	0 0 0 1 1 1 2 2 2 2 2 3 4 4 4 4 4 6
4 9	4	7 8

The top twenty home-run hitters in the Canuckland League tended to hit more home runs than the Distance League home-run hitters. This is shown by the lower leaves on the Canuckland side of the plot.

The following example shows how more than two sets of data can be compared using box-and-whisker plots.

Example 1

The following box plots compare the heights of grade 5 girls, grade 7 boys, and grade 9 students.



By comparing the box plots, determine which class of students is represented by boxes I, II, and III, respectively.

Solution:

Since the grade 5 girls are the youngest, you would expect them to be the shortest group out of the three. Therefore, box III represents the grade 5 girls.

Since the age of the grade 7 boys is in the middle of these three groups, you would expect them to be the in-between group. Therefore, box I represents the grade 7 boys.

This leaves box II for the grade 9 students. Many boys increased their height by this age. This is indicated by the longer right side of the box plot.

Do the following questions.

1. Construct a back-to-back stem-and-leaf plot for the following data. Use stems of 8, 9, 10, 11, and 12. Use the stem-and-leaf diagram to answer the following questions.

   If a player is credited with 2 RBIs (Runs Batted In), it means this player has batted in two runs with a turn at bat.

   The Top 10 RBI Hitters in the Distance League, 1991

   101 99 100 92 100 90 109 93 99 82

   The Top 10 RBI Hitters in The Canuckland League, 1991

   119 111 107 103 102 121 124 98 97 94

   a. Which stem of each league has the greatest frequency of RBI's?

   b. Which stem has the smallest frequency for each league?

   c. Which league has the better RBI hitters?

2. The following are the results from an experiment where two groups of twenty-five athletes were asked to run 100 m. Construct a box-and-whisker plot for each of the following sets of data; then answer the following questions.

Group A: Relax for two minutes, and then run 100 m.

12.81	12.65	12.04	12.51	12.20	12.28	12.36	12.68
12.67	12.38	12.66	12.06	12.52	12.49	12.02	12.96
12.53	10.97	12.64	12.47	12.94	11.74	12.17	12.52
12.35							

Group B: Relax for two minutes, take ten deep breaths, and then run 100 m.

11.81	12.19	12.05	11.89	11.90	11.91	10.44	11.79
11.46	11.78	11.44	11.61	12.04	11.92	12.63	12.33
12.07	12.03	11.48	12.05	11.58	11.17	12.30	11.90
11.72							

a. Which group has the fastest runner?

b. Which group has the higher median?

c. Which group is faster?

d. What conclusion would you derive from this data?

For solutions to **Activity 1**, turn to **Appendix A**, **Topic 4**.

Activity 2

Identify the vocabulary of probability as it applies to the likelihood of the occurrence of an event within a sample.

Many of the questions you have to answer involve a degree of uncertainty.

- Will it rain tomorrow?

  Answer: There is a 90% chance of rain tomorrow.

- Will John Iscow be re-elected?

  Answer: Based on the advance poll, John Iscow will probably be re-elected.

- Do you have high blood pressure?

  Answer: If you are over 50, your chances of having high blood pressure are much higher.

Considering the circumstances given with these answers, they may all be true. However, not one of these is a direct answer to the question.

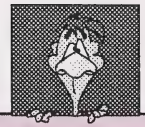
The mathematics that deals with chance or predicting is called **probability theory**. More specifically, probability predicts the outcome of an event. Examine the following case to see this.

What is the probability of rolling a two on a die?

A die has six sides and they are all equally likely to come up when the die is rolled. The number of possible outcomes is six. Only one of the faces on the die represents the number two. You can say that the number of favourable outcomes is one. Therefore, there is one chance out of six that a two will be rolled.

In probability this would be written as $P(2)=\frac{1}{6}$. This would be read as follows: The probability of rolling a two on a die is one-sixth.

This case leads us to the following probability formula.



Probability is the likelihood of something happening.

This is a die.

These are dice.

$$\text{Probability of an event } E = \frac{\text{number of favourable outcomes}}{\text{total number of possible outcomes}}$$

$$P(E) = \frac{\text{number of favourable outcomes}}{\text{total number of possible outcomes}}$$

Looking back at the die example.

$P(2) = \frac{1}{6}$

The event is that a two will show up on the die.

There is only one face on the die that is favourable.

This is the total number of possible outcomes, or the number of faces on the die.

Take a look at another case.

If a card is drawn at random from a well-shuffled deck, what is the probability it is an ace?

Identify the different components of this situation and substitute them into the formula.

The event is that an **ace** will be drawn.
There are four aces in a deck of cards.
There are fifty-two cards in a deck.

Therefore, $P(ace) = \frac{4}{52}$.

Since probability is written as a fraction, it must be written in lowest terms.

$$P(ace) = \frac{4}{52}$$
$$= \frac{4 \div 4}{52 \div 4}$$
$$= \frac{1}{13}$$

The probability of drawing an ace from a deck of cards is $\frac{1}{13}$ or 1 in 13 draws.

Continue to examine the deck of cards.

What is the probability that a selected card will have the number eleven on it?

The event is that an **eleven** will be drawn.
There are no cards with an eleven on the face in the deck of cards.
There are fifty-two cards in a deck.

Therefore, $P(eleven) = \frac{0}{52}$ or 0.

Zero is the lowest value that you can get for probability.

Getting an eleven on the face of a card is considered to be an **impossible** event.

What is the probability that a selected card will be red or black?

The event is that a **red** card or a **black** card will be drawn.
Fifty-two cards are either red or black.
There are fifty-two cards in a deck.

Therefore, $P(red\ or\ black) = \frac{52}{52}$ or 1.

One is the highest number that you can get for probability. Getting a red or black card is considered to be a **certain** event.

If you are having trouble understanding probability, turn to the **Extra Help** section at the end of this topic.

See if you can answer the following questions.

Answer any three of the following questions.

1. If one letter is selected at random from the word **dimensional**, determine the probability that it will be the following:

   a. an **m**
   b. an **i**
   c. a vowel
   d. a consonant
   e. an **o**
   f. a **d** or an **l**

2. A golf course is having a lottery to help pay for the addition of nine holes to their course. If 3000 tickets are sold, determine the probability of winning if you buy the following:

   a. 1 ticket
   b. 10 tickets
   c. 300 tickets
   d. no tickets
   e. 3000 tickets

3. Slips of paper numbered 1 to 50 are placed in a box. If one of the pieces of paper is selected at random, determine the probability that the number will be the following:

   a. a 2
   b. an even number
   c. divisible by 3
   d. both digits of the number are 3s

The word **or** indicates that one of the events must happen. The word **and** indicates that both of the events must happen.

The probability of getting heads when you toss a coin is $\frac{1}{2}$.

4. A wheel is divided into twelve equal sections that are numbered 1 to 12. If the wheel is spun, what is the probability that a pointer will be resting on the following when the wheel stops?

   a. a 5
   b. an odd number
   c. a prime number
   d. a number greater than 9

For solutions to **Activity 2**, turn to **Appendix A, Topic 4**.

Activity 3

Make a probability statement and show how it relates to a sample and the population.

Probability and statistics are closely related. Probability is often used in interpreting the results that are collected from statistics. Examine how the skills of probability are applied to answer questions in the following example.

Example 2

A survey asked 1000 people if they agreed or disagreed with the following statement.

Television drama series are too violent.

- 702 people agreed.
- 254 people disagreed.
- The rest had no comment.

Determine the probability that another person selected at random will do the following:

- agree

Solution:

$$P(agree) = \frac{702}{1000}$$

$$= \frac{351}{500} \text{ or } 0.702$$

The probability that the person selected will agree with the statement is $\frac{351}{500}$ or 0.702 or 70.2%.

- disagree

  Solution:

  $$P(disagree) = \frac{254}{1000}$$
  $$= \frac{127}{500} \text{ or } 0.254$$

  The probability that the person selected will disagree with the statement is $\frac{127}{500}$ or 0.254 or 25.4%.

- have no comment

  Solution:

  $$P(no\ comment) = \frac{44}{1000}$$
  $$= \frac{11}{250} \text{ or } 0.044$$

  The probability that the person selected will have no comment is $\frac{11}{250}$ or 0.044 or 4.4%.

Based on these results, you could say that the next person asked is likely to be in agreement with the comment.

There are additional questions that might come out of this case.

- Would you expect 70% of the people you asked to give you the same response?

- Does the no comment segment include too large a part of the sample?

Since the sample included only one thousand people, you would not expect the sample to be very accurate for predictions for the whole country. However, if the sample was taken from a town with a population of around 10 000, you would expect any prediction about the town to be quite accurate.

The no comment segment is smaller than 5% of the sample. This segment would be too small to have an influence on the findings of the survey.

Take a look at another example.

Example 3

A polling company took a poll in a provincial riding for an upcoming by-election. The polling company phoned 1000 residents in a riding that has 30 000 voters. When asked whom they would vote for, the following results were found.

NC Candidate	211
Reform Group Candidate	191
Literal Candidate	189
Old Line Candidate	167
Hippo Party Candidate	112
Not Sure	130

The polling company reported that they expected the NC Candidate to be re-elected.

Answer the following questions using the previous results.

- What percent of the sample said they would vote for the NC Candidate?

  Solution:

$$\text{NC Candidate} = \frac{211}{1000} \times 100$$
$$= 21.1\%$$

  21.1% of the sample indicated they would vote for the NC Candidate.

- What percentage of the sample said they would vote for a particular party?

  Solution:

  First, add up the votes for each party.

$$211 + 191 + 189 + 167 + 112 = 870$$

$$\text{All parties} = \frac{870}{1000} \times 100$$
$$= 87.0\%$$

  87.0% of the sample said they would vote for a particular party.

- Do you agree with the polling company's prediction?

  Solution:

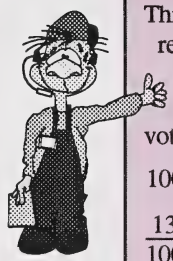
  The polling company is putting a lot of faith in the 2% lead by the NC Party. Thirteen percent of the group did not pick a particular party and may vote for any party.

  The polling company's prediction is too risky.

Are there any questions that you think should be asked and answered about this survey? How did the polling company select the residents to poll? What day of the week and what time of the day was the poll taken? Were the people who responded to the poll voters?

These are questions that could influence the outcome of the poll. Usually polling companies are aware of questions like these; thus, they make adjustments to their polling methods to make sure that such questions will not incorrectly influence a poll.

When reading predictions, ask some questions to make sure that you understand exactly what is being said and that the results are reliable.



Thirteen percent of the residents said they were not sure whom they would vote for.

$$1000 - 870 = 130$$

$$\frac{130}{1000} \times 100 = \frac{13\,000}{1000}$$
$$= 13\%$$

Answer the following questions.

1. The following information shows what percent of the people of Canuckland have a certain type of blood.

Blood Type	Percentage of People
O	45%
A	40%
B	11%
AB	4%

   a. Determine the probability that a person selected will have the following blood type.

      i. Type O

      ii. either Type A or Type B

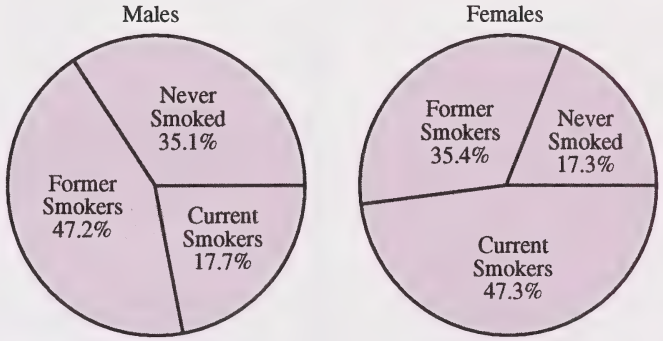
      iii. not Type B

      iv. neither Type A nor Type AB

   b. How many people would you expect to have Type O blood from a sample of 100 000 people?

   c. If a hospital keeps 40 bottles of Type AB blood on hand, how many bottles of Type A would you expect them to keep on hand?

2. The following graphs summarize data concerning smoking.



   a. What is the probability that the first man you meet on the street will be a smoker?

   b. What is the probability of a woman who is a nonsmoker to be selected randomly for a survey?

   c. Who is less likely to smoke, a man or a woman?

   d. Which group finds it easier to quit smoking?

For solutions to **Activity 3**, turn to **Appendix A, Topic 4**.

Activity 4

Comment on the confidence of predictions that are based on the analysis of a sample.

Predictions based on surveys vary in reliability, depending on factors such as the size of the sample, the manner in which the sample for the population was selected, and the nature of the questions asked. Before accepting predictions based on a sample, you should examine these factors.

To avoid difficulties with the reliability of samples, the entire population can be examined. This is called taking a **census**. The Canadian census uses the entire population. This is a reliable but an expensive and time-consuming way of gathering information.

How would you comment on the prediction based on the results of the following survey?

Forty-nine out of sixty students surveyed at Rocky Road High School brushed their teeth before leaving for school. The prediction is that approximately 82% of all high school students in Alberta brush their teeth before leaving for school.

The students at Rocky Road High School might not be representative of all the students in Alberta. The sample size is also quite small. A better prediction could be made if more students in schools throughout the province were selected. For example, you could randomly select ten high schools from a list of all schools in the province; then randomly survey 50% of the students in each school.

If you have access to a videocassette recorder, you may view the *Math Wise* video titled **Predicting – Choosing a Sample**. This program discusses what makes a good sample.

Try the following questions.

1. The following are observed occurrences that were presented as cause-and-effect relations. Is each one reasonable?

   a. A study showed that the number of flu cases increased with the number of oranges being sold in the stores.

   b. A study showed that the larger a newspaper editor's shoes, the higher his or her annual salary.

   c. A report showed that the more years of education a person has, the higher that person's salary is likely to be.

2. Would you feel confident with the following predictions? Explain your answer.

   a. A newspaper's total annual sales would increase as the number of reporters working with the newspaper increases.

   b. The Jofre 500 has an average speed of 240 km/h; therefore, the Jofre 500 should win the race on Saturday.

   c. A person's mass will increase as her or his height increases.

3. A 1983 survey of nine-year-olds found that 30% felt peer pressure from other children to insult other people. Try to write a question concerning this type of peer pressure that a nine-year-old would understand.

4. A national poll found that 81% of all parents say they have spoken with their teenagers about the dangers of drinking and driving. Only 64% of the teens say they remember such a discussion. Why do you think these results differ?

5. Eating potatoes is hazardous to your health. A survey found the following evidence to support this statement. Point out the errors that have been made to give the false conclusion that potatoes are dangerous.

   - All surviving potato eaters who were born beween 1900 and 1910 have wrinkled skin and failing hearing and/or eyesight, and they have lost most of their teeth.

   - Of those people who were the cause of a car accident, it was found that 99.8% of them had eaten potatoes.

   - Almost all of the people born in 1888, and ate potatoes, have died.

   - Most people who have died from cancer have eaten potatoes.

6. Which of the following samples will yield the most reliable prediction? The purpose of the study is to determine what percentage of students do not eat breakfast?

   a. two students out of twenty

   b. seven students out of sixty

   c. thirteen students out of one hundred

For solutions to **Activity 4**, turn to **Appendix A**, **Topic 4**.

If you require help, do the Extra Help section.

If you want more challenging explorations, do the Extensions section. } You may decide to do both.

Extra Help

Probability

Probability is a branch of mathematics concerned with the study of the chance that a given event will occur.

Since probability is a branch of mathematics, it uses numbers. The numbers that probability uses is the set of rational numbers between 0 and 1 inclusively, where 0 represents no chance of the event happening and 1 represents the fact that the event must happen. All other numbers represent degrees of the event happening.

In the previous statement, the word **event** means anything that you are looking at happening. The event might be your sports team beating another sports team, a red card being drawn from a deck of cards, or even an NC candidate being elected prime minister some time in the future. An event can be anything that might happen.

When there is no chance of an event happening or the probability of the given event is 0, this is called an **impossible event**. For example, the probability of a person being placed on Mars within the next six months is zero or impossible. You know that this is an impossible event since there are presently no plans to put a person on Mars.

When an event must happen or the probability of the given event is 1, then you have a **certain event**. An example of a certain event would be the chance of Brian Mulroney being elected to a second term to the office of Prime Minister of Canada. Again, you know that this is a certain event because it has already happened.

Probability can be defined in another way, the way it is calculated.

The probability of an event happening is the **total number of favourable outcomes divided by the total number of outcomes, where all of the outcomes have an equal chance of happening**.

Examine this statement through some examples.

Suppose you have a small bag that contains six marbles of equal size and weight. One marble is white, two marbles are red, and three marbles are blue.

When the bag is held over your head and you reach into the bag and pull out one marble, there is an equal chance of selecting any one of the marbles.

Therefore, if you consider the selection of each marble as an outcome, then all of the outcomes have an equal chance of happening. This satisfies the last part of the previous definition of probability.

> The probability of an event happening is the total number of favourable outcomes divided by the total number of outcomes, **where all of the outcomes have an equal chance of happening.**

Suppose you have been told that you won one of the marbles in the bag. You are going to be permitted to select one of the marbles from the bag. You decide that you want the white marble. The number of white marbles is the number of favourable outcomes, 1. This is stated in the middle part of the statement and is in bold type.

> The probability of an event happening is **the total number of favourable outcomes** divided by the total number of outcomes, where all of the outcomes have an equal chance of happening.

Since there are six marbles in the bag, there are a total of six possible outcomes that can happen. This is represented in the bold part of the following statement.

> The probability of an event happening is the total number of favourable outcomes divided by **the total number of outcomes**, where all of the outcomes have an equal chance of happening.

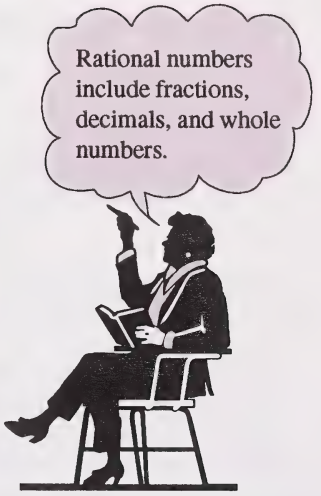
Mathematically, the probability of selecting one white marble from a bag is written as $P(w)$. In the general formula this is written as $P(E)$. The phrase that concerns that part of the case is bold in the statement that follows.

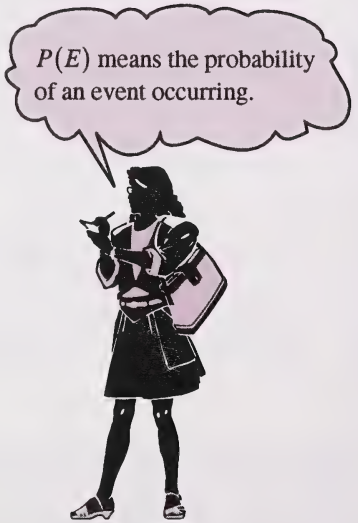
> **The probability of an event happening** is the total number of favourable outcomes divided by the total number of outcomes, where all of the outcomes have an equal chance of happening.

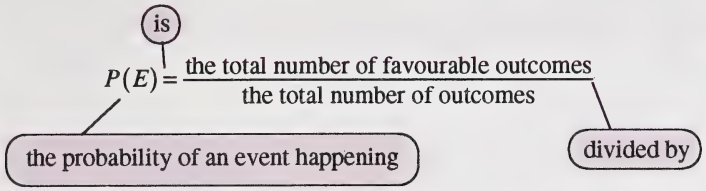
The only words that have not been highlighted are the mathematical operators; the word **is** represents **equals (=)**, and the words **divided by** represent **division (÷)**.

> The probability of an event happening **is** the total number of favourable outcomes **divided by** the total number of outcomes, where all of the outcomes have an equal chance of happening.

You have now covered all of the phrases in the statement. First, rewrite the statement into mathematical form; then, fill in the information from the example.






is
$P(E) = \frac{\text{the total number of favourable outcomes}}{\text{the total number of outcomes}}$
the probability of an event happening
divided by

$P(w) = \frac{1}{6}$

The chance of you selecting the white marble from the bag is one out of six or $\frac{1}{6}$.

Try answering the following questions.

1. A six-sided die is being thrown. Determine the probability of the following:

   a. the side with five dots will be facing up

   b. the side with one or two dots will be facing up

   The word **or** means one of the outcomes must happen. The word **and** means both of the outcomes must happen.

2. A twelve-sided die is being thrown. Determine the probability of the following:

   a. the side with seven dots will be facing up

   b. the sides with three, four, or seven dots will be facing up

   The word **or** means one of the outcomes must happen. The word **and** means both of the outcomes must happen.

3. Using a deck of 54 cards (this includes the two jokers), determine the probability of the following:

   a. a card with the number seven on the face will be drawn

   b. a red card will be drawn if you consider the jokers to be black

   c. a card with a two, a five, or a king will appear on the face of a drawn card

4. A roulette wheel has thirty-eight slots into which a marble can fall. The slots are numbered from 1 to 36 with the other two slots being labelled as 0 and 00. The 0 and 00 slots are both coloured green, while eighteen of the rest of the slots are coloured red, and the remaining eighteen slots are coloured black. Determine the probability that the marble will come to rest on the following:

   a. a green slot

   b. an even number (Do not consider the 0 and 00 as even or odd.)

   c. the number 37

For solutions to **Extra Help**, turn to **Appendix A**, **Topic 4**.

Extensions

What are the odds of winning a lottery that you play by picking six different numbers between 1 and 49?

This problem is solved most simply using the formula for combinations:

$$ {}_nC_r = \frac{n!}{r!(n-r)!} $$

In this formula n is the total number of things and r is the number of things taken at a time.

In the lottery example the number of things n is 49 since you pick from the numbers from 1 to 49, and the number of things taken at a time r is 6 (select any six numbers).

The symbol ! means factorial.

5! means $5\times4\times3\times2\times1$.
7! means $7\times6\times5\times4\times3\times2\times1$.

Put these together and see how many ways there are to pick six numbers out of 49.

$$ {}_{49}C_6 = \frac{49!}{6!(49-6)!} $$
$$ = \frac{49!}{6!43!} $$

It would be very tedious to write down all of those multiplications, but there is a trick you can use.

$$ 49! = 49\times48\times47\times46\times45\times44\times\underbrace{43\times42\times\ldots\times1}_{43!} $$

Can you see that $49! = 49\times48\times47\times46\times45\times44\times43!$?

Now write the formula again.

$$ {}_{49}C_6 = \frac{49!}{6!43!} $$
$$ = \frac{49\times48\times47\times46\times45\times44\times43!}{6\times5\times4\times3\times2\times1\times43!} $$

Cancel the 43!.

$$ \frac{43!}{43!} = \frac{\overset{1}{\cancel{43}}\times\overset{1}{\cancel{42}}\times\overset{1}{\cancel{41}}\times\ldots\overset{1}{\cancel{\times1}}}{\underset{1}{\cancel{43}}\times\underset{1}{\cancel{42}}\times\underset{1}{\cancel{41}}\times\ldots\underset{1}{\cancel{\times1}}} $$
$$ = \frac{1}{1} $$
$$ = 1 $$

You are left with the following.

$$ {}_{49}C_6 = \frac{49\times48\times47\times46\times45\times44}{6\times5\times4\times3\times2\times1} $$
$$ = \frac{10\,068\,347\,520}{720} $$
$$ = 13\,983\,816 $$

Using a calculator, follow these steps.

Enter	Display
49	49
Inv or 2nd F	49
$x!$	6.082818637^{62}
+	6.082818637^{62}
6	6
Inv or 2nd F	6
$x!$	720
+	8.448359218^{59}
43	43
Inv or 2nd F	43
$x!$	6.041526303^{52}
=	13983816

Therefore, ${}_{49}C_6 = 13\,983\,816$.

This means that there are 13 983 816 possible ways to pick six numbers out of the forty-nine available. Your chances of winning the lottery by buying one ticket are 1 out of 13 983 816. Buying 100 or even 1000 tickets does not improve your odds of winning by very much. For 100 tickets you chances would be 100 out of 13 983 816. For 1000 tickets your chances would be 1000 out of 13 983 816.

Try some of these yourself.

1. What are the odds of winning each of these lotteries if one ticket is bought in each case?

   a. a 5 - 49 lottery

   b. a 4 - 49 lottery

   c. a 3 - 49 lottery

   d. a 2 - 49 lottery

   e. a 1 - 49 lottery

   f. a 6 - 36 lottery

   g. a 5 - 36 lottery

   h. a 4 - 36 lottery

   i. a 3 - 36 lottery

   j. a 2 - 36 lottery

   k. a 1 - 36 lottery

For solutions to **Extensions**, turn to **Appendix A**, **Topic 4**.

Unit Summary

What You Have Learned

- Analyse problems and generate and collect data appropriately.
- Identify samples along with the populations from which they are taken.
- Evaluate samples for bias, appropriateness of sample type, and randomness.
- Organize data using tables, stem-and-leaf plots, and box plots.
- Graph data and interpret data graphed on pictographs, bar graphs, and circle graphs.
- Define the measures of central tendency, and determine the most suitable measure of central tendency for a given set of data.
- Determine the mean, median, and mode for discrete raw data.
- Determine the mean, median, and mode for tabular data.
- Determine the range for a given set of data.
- Compare and contrast two or more sets of data that address similar situations.
- Identify the vocabulary of probability as it applies to the likelihood of the occurrence of an event within a sample.
- Make a probability statement, and show how it relates to a sample and the population.
- Comment on the confidence of predictions that are based on the analysis of a sample.

You are now ready to

complete the **Unit Assignment**.

COURSE SURVEY FOR MATHEMATICS 10

Please evaluate this course and return this survey with your last unit assignment. This is a course designed in a new distance-learning format, so we are interested in your responses. Your constructive comments will be greatly appreciated so that a future revision may incorporate any necessary improvements.

Name _____ Course _____

Address _____ Age ❑ under 19

_____ ❑ 19 to 40

_____ ❑ over 40

File No. _____ Date _____

Design

1. This course contains a series of Unit Booklets and Assignment Booklets. Do you like the idea of separate booklets?

_____

_____

2. Have you ever enrolled in a correspondence course before that arrived as one large volume?

❑ Yes ❑ No If yes, which style do you prefer?

_____

_____

3. The Unit Booklets contained a variety of self-assessed activities. Did you find it helpful to be able to check your work and have immediate feedback?

❑ Yes ❑ No If yes, explain.

_____

_____

4. Were the questions and directions easy to understand?

❑ Yes ❑ No If no, explain.

_____

_____

5. Each topic contains follow-up activities. Which type of follow-up activity did you choose?

❑ mainly Extra Help
❑ a variety
❑ mainly Extensions
❑ none

Did you find these activities beneficial?

❑ Yes ❑ No If no, explain.

_____

_____

6. Did you understand what was expected in the Assignment Booklets?

❑ Yes ❑ No If no, explain.

_____

_____

7. The course materials were designed to be completed by students working independently at a distance. Were you always aware of what you had to do?

❑ Yes ❑ No If no, provide details.

_____

_____

Name of Student _____ Student I.D. # _____

Name of School _____ Date _____

8. Suggestions for audiocassette and videocassette activities are included in the course. Were you able to use these media options?

❑ Yes ❑ No Comment on the lines below.

_____

_____

Course Content

1. What is your overall impression of the course? Did it meet your expectations?

_____

_____

2. Did you find the work load reasonable?

❑ Yes ❑ No If no, explain.

_____

_____

3. Did you have any difficulty with the reading level?

❑ Yes ❑ No Please comment.

_____

_____

4. How would you assess your general reading level?

❑ poor reader
❑ average reader
❑ good reader

Name of Student _____ Student I.D. # _____

Name of School _____ Date _____

5. Was the material presented clearly and with sufficient depth?

❑ Yes ❑ No If no, explain.

General

1. What did you like least about the course?

2. What did you like most about the course?

Additional Comments

Name of Student _____ Student I.D. # _____

Name of School _____ Date _____

Only students enrolled with the Alberta Distance Learning Centre need to complete the remaining questions.

1. Did you contact Alberta Distance Learning Centre for help or information while doing your course?

   ❑ Yes ❑ No If yes, approximately how many times? _____

   Did you find the staff helpful?

   ❑ Yes ❑ No If no, explain.

   _____

   _____

2. Were you able to fax any of your assignments?

   ❑ Yes ❑ No If yes, comment on the value of being able to do this.

   _____

   _____

3. If you were mailing your assignments, how long was it taking for Assignment Booklets to return?

   _____

   _____

4. Was the feedback you received from your correspondence teacher helpful?

   ❑ Yes ❑ No Please comment.

   _____

   _____

   _____

   _____

Thanks for taking the time to complete this survey. Your feedback is important to us.

Fax Number: 674-6686

Instructional Design and Development Unit
Alberta Distance Learning Centre
Box 4000
Barrhead, Alberta
T0G 2P0

Name of Student _____ Student I.D. # _____

Name of School _____ Date _____

Appendices

Appendix A
Solutions

Review

Topic 1 Collection of Data

Topic 2 Organization and Presentation of Data

Topic 3 Analysis of Data: Mean, Median, Mode

Topic 4 Drawing Inferences

Appendix B
Charts and Tables

Random Number Table

Appendix A
Solutions

Review

1. Number of Blocks Travelled to School

Scale: Each ■ equals 1 block.

Similar graphs are acceptable.

2. This is an acceptable prediction since half of the tagged fish have been caught. The number 200 represents half the total number of fish in the lake.

3. The average of 7, 10, 13, 16, and 19 is $\frac{7+10+13+16+19}{5} = 13$.

4. C. line graph

5. The probability that the Blue Rockets will win their next game is $\frac{14}{16}$ or $\frac{7}{8}$.

6. The probability that the spinner will stop on a blue piece is $\frac{7}{12}$.

Exploring Topic 1

Activity 1

Analyse problems and generate and collect data appropriately.

1. Answers may vary.

   a. questionnaire, personal interview, or telephone poll of recreation centre members

   b. research records

   c. telephone poll

   d. personal interview

   e. questionnaire or personal interview

   f. experiment

2. a. sample
   b. census
   c. sample
   d. sample

Activity 2

Identify samples along with the populations from which they are taken.

1. a. 20 000 is between 10 000 and 100 000. The sample is 15% to 10% of the population.
   10% of 20 000 = 2000
   15% of 20 000 = 3000
   The sample would be at least 2000 and at most 3000.

   b. 250 is between 100 and 1000. The sample is 50% to 25% of the population.
   25% of 250 is approximately 63.
   50% of 250 = 125
   The sample would be at least 63 and at most 125.

   c. 900 000 is between 100 000 and 1 000 000. The sample is 10% to 5% of the population.
   5% of 900 000 = 45 000
   10% of 900 000 = 90 000
   The sample would be at least 45 000 and at most 90 000.

   d. 5 000 000 is between 1 000 000 and 10 000 000. The sample is 5% to 2% of the population.
   2% of 5 000 000 = 100 000
   5% of 5 000 000 = 250 000
   The sample would be at least 100 000 and at most 250 000.

2. a. Given a sample of 65, which is between 50 and 250 on the chart, the size of the population is between 100 and 1000.

   b. Given a sample of 2000, whch is between 1500 and 10 000 on the chart, the size of the population is between 10 000 and 100 000.

   c. Given a sample of 150 000, which is between 50 000 and 200 000 on the chart, the size of the population is between 1 000 000 and 10 000 000.

   d. Given a sample of 35 000, which is between 10 000 and 50 000 on the chart, the size of the population is between 100 000 and 1 000 000.

3. a. This is not a reasonable sample because the sample should be between 50 and 250.

   b. This is a reasonable sample.

   c. This is a reasonable sample.

   d. This is not a reasonable sample because the sample should be between 10 000 and 50 000.

4.

	Population Size	Sample Size	Proportion of Population
Trait 1	100 000	5000	**5%**
Trait 2	**140 000**	7000	**5%**
Trait 3	750 000	**37 500**	**5%**
Trait 4	**110 000**	5500	**5%**
Total	1 100 000	**55 000**	**5%**

Activity 3

Evaluate samples for bias, appropriateness of sample type, and randomness.

1. a. This is a convenience sample.

   b. This is a systematic sample.

   c. This is a systematic and destructive sample.

2. a. A problem in the machinery may show up only after the first thirty-five radios are produced. Problems that occur later in the day will not be discovered until the next day. Perhaps the staff and/or the machinery works better at the beginning of the day.

   b. Only the people who are interested in a new pool are interviewed. All of the people who will be involved with the cost of constructing the pool should be interviewed.

   c. This small sample will only show how well the soap will sell in Edmonton. The tests should be held across Canada to get an indication of how well it will sell in each region.

3. a. This is destructive sampling since the bulb cannot be reused.

   b. This is nondestructive sampling if the engine can be repaired.

   c. This is nondestructive sampling if tested by measuring the bounce.

   d. This is nondestructive sampling if no damage is done to the computer.

   e. This is a destructive sampling since once seed is germinated, it cannot be regerminated.

4. a. **door-to-door interviews**

Advantages	Disadvantages
A random selection can be made.	It is time-consuming and costly. Some people may refuse to cooperate.

b. **telephone poll**

Advantages	Disadvantages
A random selection can be made.	It is time-consuming and costly.
The analyst does not have to leave the office.	Some customers may refuse to cooperate.

c. **mail-in questionnaire**

Advantages	Disadvantages
It is convenient for the market analyst.	Only people with a special interest will reply.
It does not require a lot of time to collect data.	Many people will not read the questionnaire.

5. a. • Take every seventh name from a list of students. This is systematic sampling. Note that 1 out of 7 is approximately 15%.

   • Take 15% of the students in each grade. This is stratified sampling.

   • Take the first 750 students you meet in the hall. This is convenience sampling.

   b. Pick every fiftieth name in the phone book or from a city census. This is systematic sampling.

   c. • Send out a survey from a motorcycle magazine. This is self-selected sampling.

   • Take every hundredth person who has bought a motorcycle license plate. This is systematic and stratified sampling.

Extra Help

1. The average should be between 4500 and 6000.

   Suppose you choose rows 1 and 5. All the four-digit numbers when added will be as follows:

   $3570 + 7996 + 3150 + 3147 + 4845 + 8546 + 6771 + 6013$
   $+ 3969 + 7957 + 3862 + 3969 + 7797 + 6230 + 3970 + 9583$
   $+ 8145 + 9028 + 7915 + 6352 + 0650 + 4956 + 4774 + 4854$
   $+ 7450 = 145\,499$

   The average is $145\,499 \div 25 = 5819.96$ which is between 4500 and 6000.

2. The average should be between 450 and 550.

   Suppose you choose the seventh column. The first three-digit numbers from each group of five digits when added will be as follows:

   $133+157+576+731+065+673+059+624+488+648$
   $+202+854+212+577+833+134+124+750+070+200$
   $+753+412+696+331+387+966+982+731+253+458$
   $+065+983+547+031+250+458+997+994+868+448$
   $+397+092+449+659+174+446+358+678+741+060$
   $=23\,774$

   The average is $23\,774 \div 50 = 475.48$ which is between 450 and 550.

3. There are 500 000 discs.

   a. The percentage of the discs to be tested is 0.1%.

      $0.001 \times 500\,000 = 500$

      There will be 500 discs tested.

   b. The discs will be numbered from 1 to 500 000.
      The number 500 000 has six digits.
      Six digits will be required to represent each disc.

   c. Each number will have six digits. You need to select 500 numbers.

      $6 \times 500 = 3000$

      You need 3000 digits. There are fifty digits in each row since each row contains ten blocks of five numbers each.

      $3000 \div 50 = 60$

      You would require sixty rows.

Extensions

1. Machine 1 will stamp the CDs for Dadonna, U3$\frac{1}{2}$, and Von Jodi.

   These are the CDs produced for each artist.

Dadonna	400 000
U3$\frac{1}{2}$	750 000
Von Jodi	350 000
Total	1 500 000

   There will be 1 500 000 CDs produced on Machine 1.

   Machine 2 will stamp the CDs for Belinda Aisle and Neil Zirconia.

These are the CDs produced for each artist.

Belinda Aisle	300 000
Neil Zirconia	175 000
Total	475 000

There will be 475 000 CDs produced on Machine 2.

Machine 3 will stamp the CDs for Brian Odums, Flatwood Joe, and Men Out of Work.

These are the CDs produced for each artist.

Brian Odums	150 000
Flatwood Joe	250 000
Men Out of Work	125 000
Total	525 000

There will be 525 000 CDs produced on Machine 3.

2. Machine 1 will produce 1 500 000 discs. Machine 2 and will produce 475 000 discs each. Machine 3 will produce 525 000 discs.

Machine 1:

$$1\,500\,000 \text{ discs} \div \left(62\tfrac{1}{2} \times \frac{16 \text{ h}}{\text{day}}\right) = 1\,500\,000 \times \frac{1}{62\frac{1}{2}} \times \frac{\text{day}}{16 \text{ h}}$$

$$= 1500 \text{ discs/h}$$

Machine 1 will produce 1500 discs per hour.

Machine 2:

$$475\,000 \text{ discs} \div \left(62\tfrac{1}{2} \times \frac{16 \text{ h}}{\text{day}}\right) = 475\,000 \times \frac{1}{62\frac{1}{2}} \times \frac{\text{day}}{16 \text{ h}}$$

$$= 475 \text{ discs/h}$$

Machine 2 will produce 475 discs per hour.

Machine 3:

$$525\,000 \text{ discs} \div \left(62\tfrac{1}{2} \times \frac{16 \text{ h}}{\text{day}}\right) = 525\,000 \times \frac{1}{62\frac{1}{2}} \times \frac{\text{day}}{16 \text{ h}}$$

$$= 525 \text{ discs/h}$$

Machine 3 will produce 525 discs per hour.

The three machines put together will produce
$1500 + 475 + 525 = 2500$ discs in one hour.

3. Machine 1 produces 1500 discs per hour. Machine 2 produces 475 discs per hour. Machine 3 produces 525 discs per hour. Three percent of the discs will be checked.

Machine 1
3% of 1500
$0.03 \times 1500 = 45$
Machine 1 should have 45 discs checked per hour.

Machine 2
3% of 475
$0.03 \times 475 = 14.25$ or 15
(Round a part CD up to the next whole number.)
Machine 2 should have fifteen discs checked per hour.

Machine 3
3% of 525
$0.03 \times 525 = 15.75$ or 16
Machine three should have sixteen discs checked per hour.

4. Dadonna $400\,000 \times 0.01 = 4000$
   U3$\frac{1}{2}$ $750\,000 \times 0.01 = 7500$
   Von Jodi $350\,000 \times 0.01 = 3500$
   Machine 1 15 000

   Belinda Aisle $300\,000 \times 0.01 = 3000$
   Neil Zirconia $175\,000 \times 0.01 = 1750$
   Machine 2 4750

   Brian Odums $150\,000 \times 0.01 = 1500$
   Flatwood Joe $250\,000 \times 0.01 = 2500$
   Men Out of Work $125\,000 \times 0.01 = 1250$
   Machine 3 5250

5. Machine 1 produces 1500 discs per hour. Machine 2 produces 475 discs per hour. Machine 3 produces 525 discs per hour. One percent of the discs will be checked.

Machine 1
1% of 1500
$0.01 \times 1500 = 15$
Machine 1 should have fifteen discs checked per hour.

Machine 2
1% of 475
$0.01 \times 475 = 4.75$ or 5
Machine 2 should have five discs checked per hour.

Machine 3
1% of 525
$0.01 \times 525 = 5.25$ or 6
Machine 3 should have six discs checked per hour.

Exploring Topic 2

Activity 1

Organize data using tables, stem-and-leaf plots, and box plots.

1. First arrange the numbers in ascending order of value.

   309, 311, 316, 327, 334, 338, 344, 352, 352, 352, 354, 356, 357, 359, 361, 362, 363, 364, 364, 364, 378, 382, 383, 385

a. Stem-and-Leaf Plot

The Depth of Holes in an Oil Field
(in metres)

Stem	Leaf
30	9
31	1 6
32	7
33	4 8
34	4
35	2 2 2 4 6 7 9
36	1 2 3 4 4 4
37	8
38	2 3 5

b. Box-and-Whisker Plot

309	311	316	327	334	338
344	352	352	352	354	356
357	359	361	362	363	364
364	364	378	382	383	385

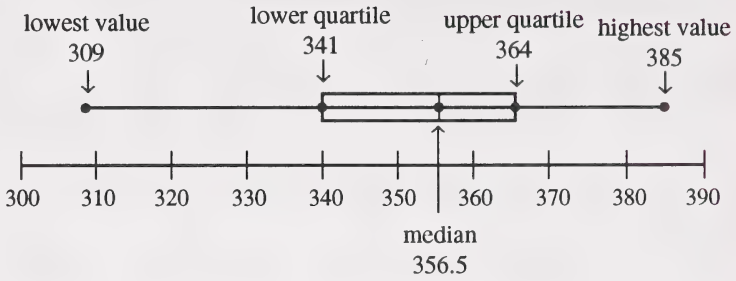
$$\text{Median} = \frac{356 + 357}{2}$$
$$= 356.5$$

$$\text{Lower quartile} = \frac{338 + 344}{2}$$
$$= 341$$

$$\text{Upper quartile} = \frac{364 + 364}{2}$$
$$= 364$$

The extremes are 309 and 385.

The Depth of Holes in an Oil Field
(in metres)



2. First arrange the data in ascending order of value.

1024, 1064, 1098, 1143, 1266, 1267, 1288, 1356, 1434, 1436, 1497, 1498, 1544, 1548, 1554, 1603, 1604, 1730, 1732, 1739, 1840, 1872, 1876, 1920

a. Stem-and-Leaf Plot

Population in Small Alberta Towns

Stem	Leaf
10	24 64 98
11	43
12	66 67 88
13	56
14	34 36 97 98
15	44 48 54
16	03 04
17	30 32 39
18	40 72 76
19	20

b. Box-and-Whisker Plot

1024	1064	1098	1143	1266	1267
1288	1356	1434	1436	1497	1498
1544	1548	1554	1603	1604	1730
1732	1739	1840	1872	1876	1920

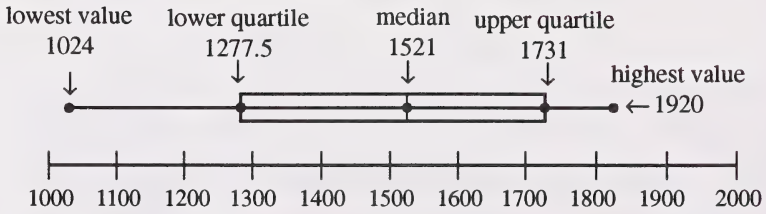
$$\text{Median} = \frac{1498+1544}{2}$$
$$= 1521$$

$$\text{Lower quartile} = \frac{1267+1288}{2}$$
$$= 1277.5$$

$$\text{Upper quartile} = \frac{1730+1732}{2}$$
$$= 1731$$

The extremes are 1024 and 1920.

Population in Small Alberta Towns



3. Percentage of Fast Food Sold

Type of Food	Frequency	Percentage
Coffee	81	16%
Hot dogs	26	5%
French fries	98	20%
Hamburgers	134	27%
Onion rings	29	6%
Pop	132	26%
Total	500	100%

The percentages have been rounded to the nearest percent.

4. Number of Innings Thrown

Pitcher	Innings Thrown	Percentage
Dodd	90	18%
Foggson	72	14%
Freelance	123	25%
Garcia	67	13%
Inrate	23	5%
Morris	90	18%
Stanton	12	2%
Worson	20	4%
Zielert	3	1%
Total	500	100%

5. How Different Types of Music Sells
(Survey Conducted in the A&S Record Store on February 2, 1992)

Type of Music	Frequency	Percentage of Total
Blues	12	6%
Classical	32	16%
Country	34	17%
Reggae	6	3%
Soul	42	21%
Rock	74	37%
Total	200	100%

a. The most popular music is rock.

b. The least popular music is reggae.

c. The survey included 200 people.

d. Seventeen percent of the people preferred country music.

e. The survey was done in the A&S Record store on February 2, 1992.

f. A total of 54 people (12 for blues and 42 for soul) prefer blues or soul, and 74 people prefer rock. Therefore, more people prefer rock than those preferring blues and soul combined.

6.

Stem	Leaf
0	0 2 3 8 9
1	0 1 1 5 7 7 8
2	0 2 4 4 7 8 8 9
3	2 5 6 8
4	3
5	2

a. There are twenty-six players on the Bulls team.

b. Ten players hit more than twenty-five home runs.

c. The Bulls played 120 games.

d. The player who hit fifty-two home runs hit the most home runs.

e. The person who hit no home runs or zero home runs has the least number of home runs.

f. The exact number of home runs hit by players who hit between 30 and 39 home runs is 32, 35, 36, and 38.

7. a. The shortest student has a height of about 148 cm.

   b. The tallest student has a height of about 212 cm.

   c. The median height of the students in the school is about 183 cm.

   d. One-half are below the median or between 148 cm and 183 cm. One-half are above the median or between 183 cm and 212 cm. One-half are in the box or between 172 cm and 193 cm. One-half are in the two whiskers or between 148 cm and 172 cm and between 193 cm and 212 cm.

   e. All measurements are in centimetres.

   f. Most of the heights of the students are fairly evenly distributed with a small clustering near the centre of the measures.

Activity 2

Graph data and interpret data graphed on pictographs, bar graphs, and circle graphs.

1. a. The graphs illustrate the changing population distribution in Adanacland in the years 1901, 1941, and 1986.

   b. The percentages in each of the three graphs do not add up to 100%. Because the total percentage is within 2% of 100%, these totals are acceptable.

   c. The Eastland and Francois areas decreased through the three graphs.

   d. The population for Centreland is as follows:
   - In 1901 it was 40% of 5 371 000 which is 2 148 000.
   - In 1941 it was 33% of 11 507 000 which is 3 797 000.
   - In 1986 it was 36% of 25 384 000 which is 9 138 000.

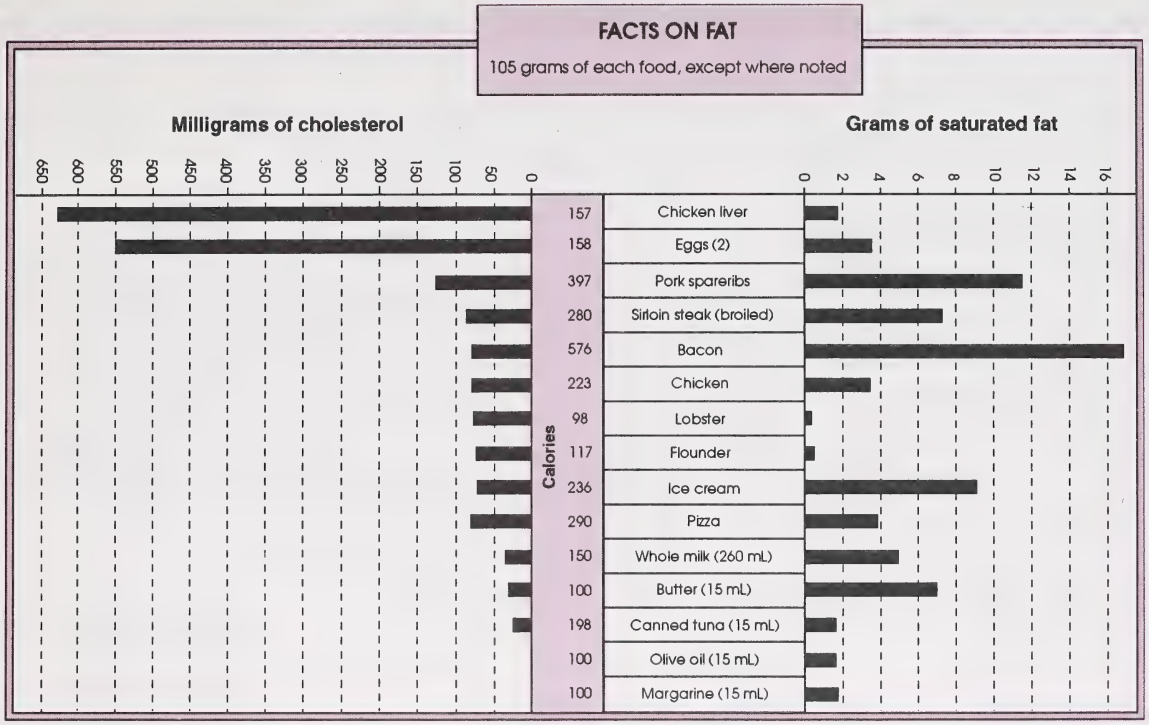
   e. The population of Eastland decreased the most from 1901 to 1986.

   f. The total population for Adanacland is as follows:
   - In 1901 it was 5 371 000.
   - In 1941 it was 11 507 000.
   - In 1986 it was 25 384 000.

   g. Great Plains has increased from 9% to 18% of the total population.

   h. In 1901 the population of Great Plains was 9% of 5 371 000 which is 483 390 and in 1986 it was 18% of 25 384 000 which is 4 569 120. The population has increased by $4\,569\,120 \div 483\,390$ or 9.5 times even though the percentage has only doubled.

2.


FACTS ON FAT
105 grams of each food, except where noted
Milligrams of cholesterol
Grams of saturated fat
650 600 550 500 450 400 350 300 250 200 150 100 50 0
0 2 4 6 8 10 12 14 16
Calories
157 Chicken liver
158 Eggs (2)
397 Pork spareribs
280 Sirloin steak (broiled)
576 Bacon
223 Chicken
98 Lobster
117 Flounder
236 Ice cream
290 Pizza
150 Whole milk (260 mL)
100 Butter (15 mL)
198 Canned tuna (15 mL)
100 Olive oil (15 mL)
100 Margarine (15 mL)

a. The quantities of food being used are 105 g, 260 mL, 15 mL, and 2 eggs.

b. Olive oil and margarine both show no cholesterol on the graph.

c. Lobster has the least amount of saturated fat. Lobster has between about 0.2 g and 0.4 g of saturated fat.

d. Chicken has 223 calories in a 100 g serving.

e. The two scales show the amount of cholesterol and the amount of saturated fat.

f.
- The food which has the most calories is bacon.
- The food with the most cholesterol is chicken liver.
- The food with the most saturated fat is bacon.

g. One litre of milk will have between 135 mg and 150 mg of cholesterol.

$$\frac{37 \text{ mg}}{260 \text{ mL}} = \frac{x}{1000 \text{ mL}}$$

$$260x = 37\,000$$

$$\frac{260x}{260} = \frac{37\,000}{260}$$

$$x \doteq 142.307\,692\,3$$

$$\doteq 142 \text{ mg}$$

h. Bacon has the highest combination in two of the categories (calories and saturated fat).

3. The Number of Employees for Canuckland's Seven Largest Employers

Each symbol represents 10 000 people.

a. Each shape represents 10 000 people.

b. The Eddie Watson Company has $6.7 \times 10\,000$ or about 67 000 employees. (Note that 66 000 to 68 000 is acceptable.)

c. CPS has $6.5 \times 10\,000$ or about 65 000 employees. Southern Fried Fish has $5.5 \times 10\,000$ or about 55 000 employees.

CPS has about 10 000 more employees than Southern Fried Fish.

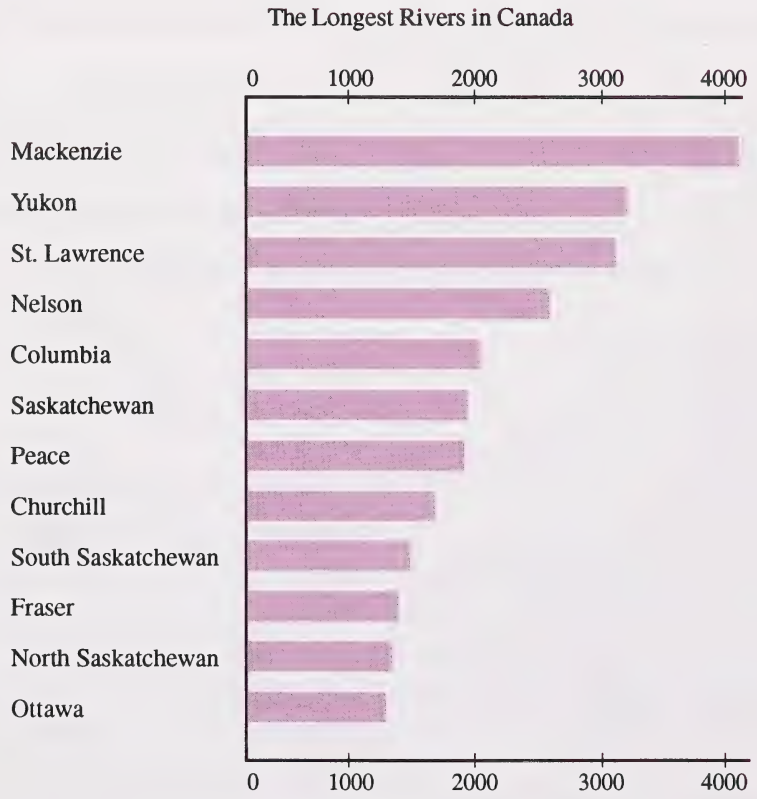
d. There are $53.7 \times 10\,000$ or about 537 000 employees represented on the graph. (Note that 536 000 to 538 000 is acceptable.)

4. The Number of Licensed Automobile Owners in Canada in 1991

Legend: Each ● represents 100 000 automobile owners.

Your graph may appear slightly different than this graph.

5.



Source: Canada Year Book 1988, Statistics Canada[1]

Your graph may appear slightly different than this graph.

[1] Statistics Canada. *Canada Yearbook*. Reprinted with permission of the Minister of Supply and Sevices Canada.

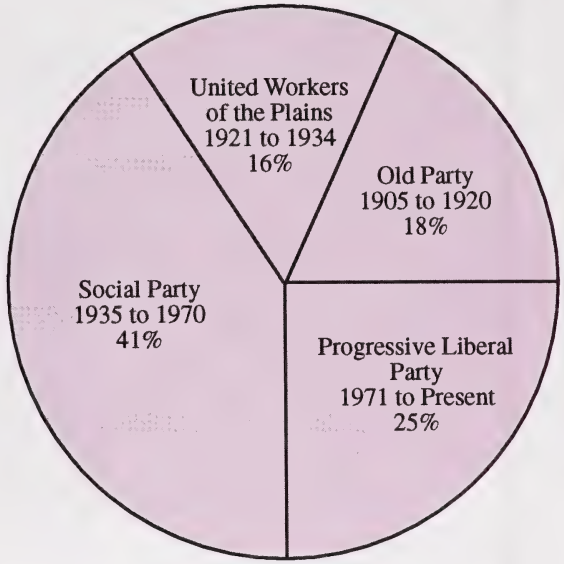
6.

Party	Years
Old Party	16
United Workers of the Plains	14
Social Party	36
Progressive Liberal Party	22*
Total	88

*The year 1992 was used for this table.

Percentage	Degrees
$\frac{16}{88} \times 100\% \doteq 18.181\,818\,18\%$ $\doteq 18\%$	$0.18 \times 360 = 64.8°$ $\doteq 65°$
$\frac{14}{88} \times 100\% \doteq 15.909\,090\,91\%$ $\doteq 16\%$	$0.16 \times 360 = 57.6°$ $\doteq 58°$
$\frac{36}{88} \times 100\% \doteq 40.909\,090\,91\%$ $\doteq 41\%$	$0.41 \times 360 = 147.6°$ $\doteq 148°$
$\frac{22}{88} \times 100\% = 25\%$	$0.25 \times 360 = 90°$
Totals: 100%	361°

The Governing Parties of the Plains (1905 to 1992)


United Workers of the Plains 1921 to 1934 16%
Old Party 1905 to 1920 18%
Social Party 1935 to 1970 41%
Progressive Liberal Party 1971 to Present 25%

Your graph may appear slightly different than this graph.

7. The pictograph and the bar graph are both used when you want to make a visual comparison of similar data which is grouped into separate and distinct categories. However, the bar graph is used when you want more accuracy and the pictograph is used when you want to be more appealing.

   A circle graph, on the other hand, is used when you want to compare the separate parts to the whole.

Activity 3

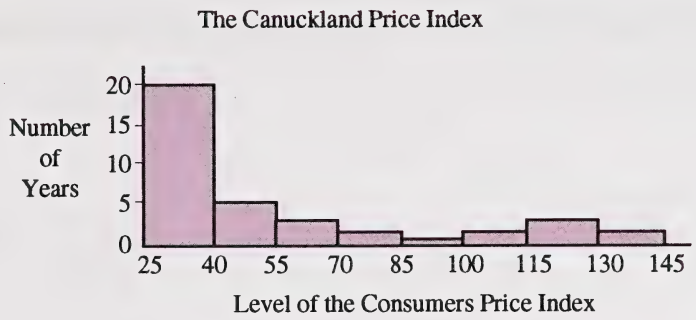
Graph and interpret continuous data using histograms and frequency polygons.

1. Your graph may be slightly different.

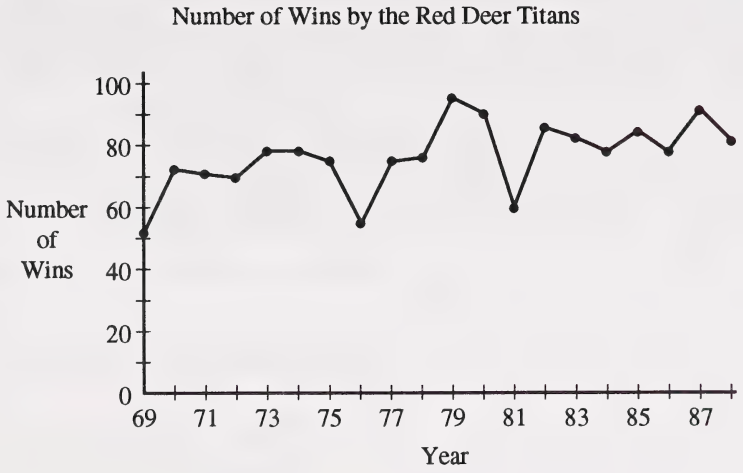
$$\text{Range} = 138 - 25$$
$$= 113$$

$$\text{Class width} = \frac{113}{8}$$
$$= 14.125 \text{ rounded to } 15$$

Class	Class Interval	Tally	Frequency
1	25-39	𝍸 𝍸 𝍸 𝍸	20
2	40-54	𝍸	5
3	55-69	III	3
4	70-84	II	2
5	85-99	I	1
6	100-114	II	2
7	115-129	III	3
8	130-144	II	2



2. Your graph may be slightly different.



3. a. The most number of wins that is possible on the graph is 119 or 120.

   b. Class 4 has the greatest number of teams.

   c. The 105 to 109 interval has seven teams.

   d. Twenty-two teams won between 90 and 94 games.

   e. Sixty-eight teams or sixty-eight years are used in this histogram.

4. a. The population in 1911 was about seven million.

   b. The population in 1931 was 10 300 000.

   c. The greatest increase in the population was between 1951 and 1961.

   d. The data was most likely collected through a census by the Government of Canuckland.

   e. The population reached twenty million between 1961 and 1971.

5. Histograms based on the same data will look completely different when the number of classes used in each histogram is completely different. Histograms should always be made with about eight to ten classes.

Extra Help

1.

Ages	Tally	Frequency
14	IIII	4
15	𝍸 I	6
16	𝍸 IIII	9
17	𝍸 I	6
18	IIII	4
19	I	1

The total of the survey frequencies is 30.

2. a. There are nine students that are sixteen years old.

   b. There were $9 - 6 = 3$ more sixteen-years-old students than fifteen-years-old students.

   c. $\frac{9}{30} \times 100\% = 30\%$
   Thirty percent of the students surveyed are sixteen years old.

Extensions

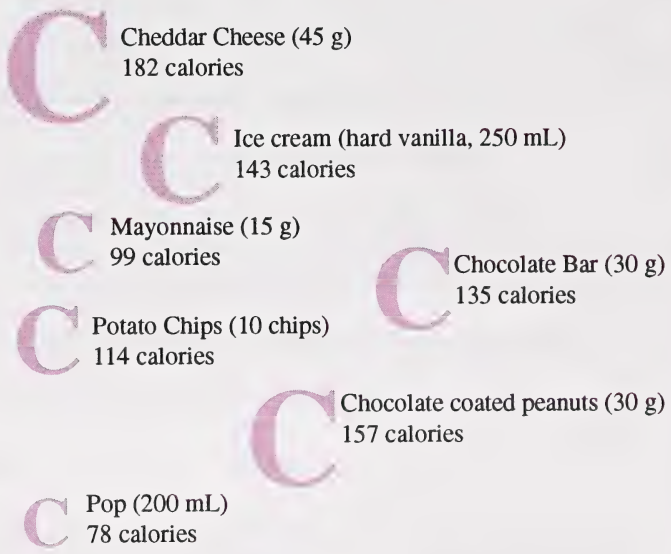
1. $U3\frac{1}{2}$ has the largest picture.

2. The smallest amount is 125 000 discs.

3. Every 1 mm on the graph represents 12 000 discs.

4. The $U3\frac{1}{2}$ circle is about four times the Neil Zirconia circle since $\frac{63 \text{ mm}}{15 \text{ mm}} = 4.2$.

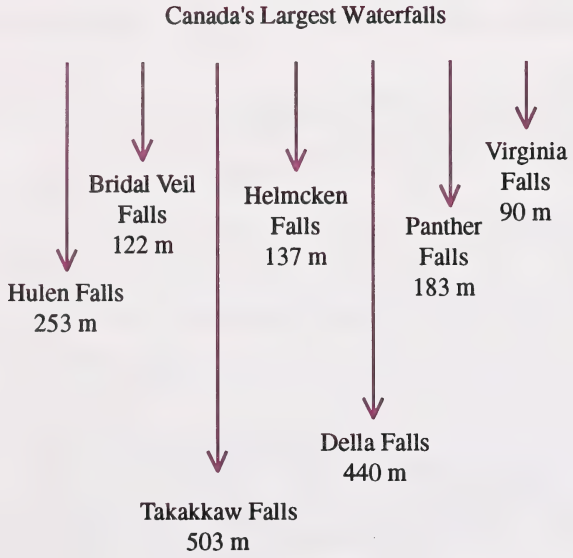
5. The Amount of Energy in Food



Scale: 1 mm in height = 10 calories

You may have used a different picture or symbol for this graph.

6.



Source: Canada Year Book 1988, Statistics Canada[1]
Scale: 1 mm = 8 m

You may have used a different picture for this graph.

[1] Statistics Canada. *Canada Yearbook*. Reprinted with permission of the Minister of Supply and Services Canada.

Exploring Topic 3

Activity 1

Define the measures of central tendency, and determine the most suitable measure of central tendency for a given set of data.

1. a. Use the mean for a weekly average salary.

   b. Use the mean for a curling average.

   c. Use the mean for the amount of rainfall a town receives.

   d. Use the median for the average mark on a test in a class.

   e. Use the mean for the cost of gasoline in a town.

   f. Use the mode for the most popular dress size.

   g. Use the mode for the number of crackers in a box.

   h. Use the median or mean for the number of goals scored per game in soccer.

   i. Use the median or mean for the number of hours spent studying each night.

j. Use the median or mean for the number of hours spent watching TV each night.

k. Use the mode for the average number of people per family.

2. It is not necessary to put the data in ascending order before finding the mean. To find the mean, add the numbers and divide by the number of data.

3. It is not necessary, but it will make your work considerably easier if the data is organized before searching for the mode.

4. When finding the median, first arrange the data in ascending order.

Activity 2

Determine the mean, median, and mode for discrete raw data.

1. The mean is as follows:

$$\bar{x} = \frac{35+34+39+35+30+34+32+38+34+39}{9}$$

$$= \frac{350}{10}$$

$$= 35$$

The median is as follows:

First arrange the data in ascending order.

30 32 34 34 34 35 35 38 39 39

The median is the average of 34 and 35.

$$\text{Median} = \frac{34+35}{2}$$

$$= \frac{69}{2}$$

$$= 34.5$$

The mode is as follows:

The value 34 occurs most frequently in the data (three times) and therefore, it is the mode.

2. The median will be the best measure for an average since one of the values is an extreme.

$1700 $2100 $2100 $2100 ($2400) $2500
$2500 $3200 $5200

The middle value is $2400. Therefore, the average monthly wage is $2400.

3. The median will be the best measure for an average since two of the values are extremes.

9 10 16 19 (24) 26 154 1416 5834

The middle value is 24. Therefore, the average length of one day in the solar system is twenty-four hours.

4. The mean will provide the best average for the data since there are no extreme values.

$$\bar{x} = \frac{5025}{21}$$
$$\doteq 239.3$$

The mean is 239.3 g.

5. The mode is the best measure for this data since the average should be one of the actual values.

165 mm 170 mm 175 mm 175 mm 180 mm 180 mm 180 mm

185 mm 185 mm 185 mm 185 mm 185 mm 185 mm 185 mm

185 mm 190 mm 190 mm 195 mm 195 mm 195 mm 200 mm

There are eight measures of 185 mm; therefore, the mode is 185 mm, and the average of the hat sizes is 185 mm.

Activity 3

Determine the mean, median, and mode for tabular data.

1. The Top 30 at Wolf Greens

27	6 8
28	0 0 3 6 6 9
29	0 1 1 1 5 5 7 9 9
30	2 2 5 5 8 8 8
31	0 0 6 7 7 9

The mean is as follows:

$$\bar{x} = \frac{276+278+280+280+\ldots+319}{30}$$
$$= \frac{8933}{30}$$
$$\doteq 297.766\,667$$
$$\doteq 297.8$$

The mean is approximately 297.8 strokes when rounded to the nearest tenth.

The median is shown with a rectangle on the plot.

$$\text{Median} = \frac{297+299}{2}$$
$$= 298$$

The median is 298 strokes.

The mode is shown with an ellipse on the plot.

This particular group of data is bimodal. The two modes are 291 and 308.

2. The Points Earned by the Banff Boomers, 1955 to 1990

7	5 5 9
8	1 2 3 5 7 7
9	0 1 (2 2 2 2) 3 4 [6 7] 8 8 9
10	0 3 3 3 4 7 9
11	3 5 8
12	0 7 9
13	2

The mean is as follows:

$$\bar{x} = \frac{75+75+79+81+82+\ldots+132}{36}$$

$$= \frac{3541}{36}$$

$$\doteq 98.36111111$$

$$\doteq 98.4$$

The mean is about 98.4 points when rounded to the nearest tenth.

The median is shown with a rectangle on the plot.

$$\text{Median} = \frac{96+97}{2}$$

$$= 96.5$$

The median is 96.5 points.

The mode is shown with an ellipse on the plot.

There is only one mode. The mode is 92 points since the one's digit 2, appears most frequently in the separate rows.

3. a. The median is about 152.

b. The median is about 191.

c. The median is about 150.

d. The median is about 172.

e. The median is about 162.

f. The median is about 211.

Activity 4

Determine the range for a given set of data.

1. a. Points Scored in the 1988/1989 DBA Season

(2868)	1938	1669	2397	1932	1621	1704
2275	2000	1607	2264	1641	1653	1674
1574	1446	(1446)	1805	1938	2185	2268

$$\text{Range} = 2868 - 1446$$

$$= 1422$$

The range is 1422 points.

b. Goals Scored by the Scoring Champions (DHL)

2	(3) 6 6 8 9
3	0 1 3 3 3 5 6 6 8 8 9 9
4	0 1 3 4 6 7 7 7 9 9
5	0 0 2 3 4 5 5 6 6
6	0 2 6 8
7	0 1 3 6
8	7
9	(2)

Range = 92 – 23

= 69

The range is 69 goals.

2. a. Wolf Greens Amateur Championship Scores in 1987

Range = 319 – 256

= 63

The range is 63 strokes.

b. Wolf Greens Amateur Championship Scores in 1988

Range = 320 – 260

= 60

The range is 60 strokes.

c. Wolf Greens Amateur Championship Scores in 1989

Range = 316 – 258

= 58

The range is 58 strokes.

Extra Help

1. a. For 23, there is an odd number of values; thus, Method 1 will be used.

b. For 105, there is an odd number of values; thus, Method 1 will be used.

c. For 56, there is an even number of values; thus, Method 2 will be used.

d. For 186, there is an even number of values; thus, Method 2 will be used.

2. a. 23, 65, 39, 92, 28, 63, 27, 83, 65, 73

There are ten values. Therefore, Step 3 from Method 2 will be used.

$\frac{10}{2} = 5$

The number will be 5.

b. 184, 458, 382, 826, 382, 572, 865, 184, 34

There are nine values. Therefore, Step 3 from Method 1 will be used.

$\frac{9}{2} = 4.5$

$= 5$

The number will be 5.

c. 12, 382, 7783, 239, 22, 264, 2002, 49, 340, 67, 49

There are eleven values. Therefore, Step 3 from Method 1 will be used.

$\frac{11}{2} = 5.5$

$= 6$

The number will be 6.

3. a. 8, 6, 4, 7, 3, 9, 6, 3, 34, 19, 4, 15, 8

First rearrange the data into ascending order.
3, 3, 4, 4, 6, 6, 7, 8, 8, 9, 15, 19, 34

There is an odd number of values. Use Method 1 since the number is 13.

$\frac{13}{2} = 6.5$

$= 7$ (Use the seventh value from either end.)

3, 3, 4, 4, 6, 6, (7), 8, 8, 9, 15, 19, 34

Seven values to middle — Seven values to middle

The median is 7.

b. 12, 27, 24, 18, 20, 22, 13, 16, 15, 23, 45, 1

First rearrange the values into ascending order.
1, 12, 13, 15, 16, 18, 20, 22, 23, 24, 27, 45

There is an even number of values. Use Method 2 since the number is 12.

$\frac{12}{2} = 6$

Use the sixth value from either end.

1, 12, 13, 15, 16, (18, 20), 22, 23, 24, 27, 45

Six values to middle — Six values to middle

The median is $\frac{18+20}{2} = 19$.

c. 102, 116, 89, 115, 121, 98, 92, 107, 113, 114, 97

First rearrange the data into ascending order.
89, 92, 97, 98, 102, 107, 113, 114, 115, 116, 121

There is an odd number of values. Use Method 1 since the number is 11.

$\frac{11}{2} = 5.5$

$= 6$ (Use the sixth value from either end.)

89, 92, 97, 98, 102, (107), 113, 114, 115, 116, 121

Six values to middle — Six values to middle

The median is 107.

Extensions

1. Number of Vehicles Produced per Month in 1989 (in 100 000's)

	Tonkyo				Nosson				Sozokki			
	x	$\bar{x}$	dif	dev	x	$\bar{x}$	dif	dev	x	$\bar{x}$	dif	dev
January	17	17	0	0	13	17	−4	4	3	17	−14	14
February	17	17	0	0	23	17	6	6	20	17	3	3
March	17	17	0	0	15	17	−2	2	19	17	2	2
April	17	17	0	0	21	17	4	4	22	17	5	5
May	17	17	0	0	14	17	−3	3	21	17	4	4
June	17	17	0	0	16	17	−1	1	17	17	0	0
	Total			0	Total			20	Total			28

$$\bar{M} = \frac{0}{6} \qquad \bar{M} = \frac{20}{6} \qquad \bar{M} = \frac{28}{6}$$

$$= 0 \qquad \doteq 3.3 \qquad \doteq 4.7$$

Range: $17 - 17 = 0$ $\qquad 23 - 13 = 10$ $\qquad 22 - 3 = 19$

The mean deviation downplays the difference in the values.

2. a. Number of Hits by the Top Six Players in 1989

Red Deer				St. Paul				Jasper			
x	$\bar{x}$	dif	dev	x	$\bar{x}$	dif	dev	x	$\bar{x}$	dif	dev
184	138	46	46	164	144	20	20	179	162	17	17
164	138	26	26	160	144	16	16	178	162	16	16
152	138	14	14	155	144	11	11	163	162	1	1
116	138	−22	22	133	144	−11	11	163	162	1	1
114	138	−24	24	128	144	−16	16	144	162	−18	18
100	138	−38	38	125	144	−19	19	143	162	−19	19
Total			170	Total			93	Total			72

$$\bar{M} = \frac{170}{6} \qquad \bar{M} = \frac{93}{6} \qquad \bar{M} = \frac{72}{6}$$

$$\doteq 28.3 \qquad = 15.5 \qquad = 12$$

b. The Jasper team has the smallest mean deviation.

3. a. Average Temperature for Selected Canadian Cities

	Red Deer				Quebec				Vancouver			
	x	$\bar{x}$	dif	dev	x	$\bar{x}$	dif	dev	x	$\bar{x}$	dif	dev
January	−12	1	−13	13	−10	3	−13	13	3	8	−5	5
February	−7	1	−8	8	−9	3	−12	12	5	8	−3	3
March	−4	1	−5	5	−3	3	−6	6	6	8	−2	2
April	3	1	2	2	6	3	3	3	8	8	0	0
May	9	1	8	8	13	3	10	10	12	8	4	4
June	14	1	13	13	18	3	15	15	14	8	6	6
		Total		49		Total		59		Total		20

$\overline{M} = \frac{49}{6}$ $\doteq 8.2$

$\overline{M} = \frac{59}{6}$ $\doteq 9.8$

$\overline{M} = \frac{20}{6}$ $\doteq 3.3$

The answers to questions 2 and 3 are rounded to one decimal place.

Note: The mean temperature ($\bar{x}$) has been rounded to the nearest whole number for the first two cities.

b. Quebec has the largest mean deviation.

Exploring Topic 4

Activity 1

Compare and contrast two or more sets of data that address similar situations.

1. Top 10 RBI Hitters in the Distance and Canuckland Leagues

Distance League		Canuckland League
2	8	
0 2 3 9 9	9	4 7 8
0 0 1 9	10	2 3 7
	11	1 9
	12	1 4

a. The Distance League's longest stem is the 90 stem and the Canuckland League's longest stems are the 90 and 100 stems.

b. The Distance League has two stems that have no leaves: 110 and 120 stems. The 80 stem for the Canuckland League also has no leaves.

c. The Canuckland League has the better RBI hitters.

2. Group A:

10.97	11.74	12.02	12.04	12.06	(12.17)
(12.20)	12.28	12.35	12.36	12.38	12.47
12.51	12.52	12.52	12.53	12.64	(12.65)
(12.66)	12.67	12.68	12.81	12.94	12.96

12.49

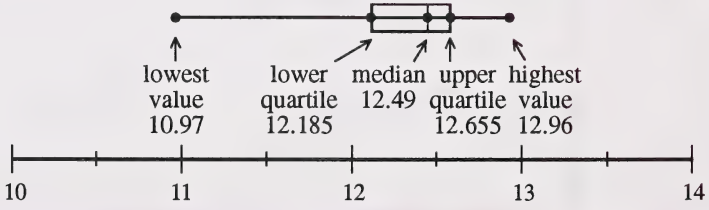
The median is 12.49.

$$\text{Lower quartile} = \frac{12.17+12.20}{2}$$
$$= \frac{24.37}{2}$$
$$= 12.185$$

$$\text{Upper quartile} = \frac{12.65+12.66}{2}$$
$$= \frac{25.31}{2}$$
$$= 12.655$$

The lowest value is 10.97.

The highest value is 12.96.



Group B:

10.44	11.71	11.44	11.46	11.48	(11.58)
(11.61)	11.72	11.78	11.79	11.81	11.89
11.90	11.91	11.92	12.03	12.04	(12.05)
(12.05)	12.07	12.19	12.30	12.33	12.63

11.90

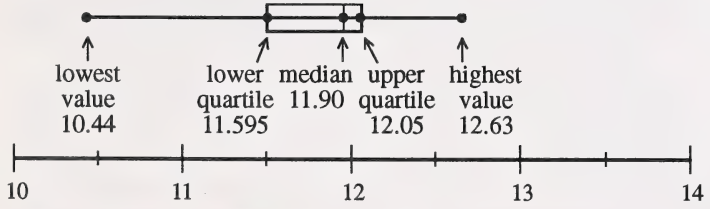
The median is 11.90.

$$\text{Lower quartile} = \frac{11.58+11.61}{2}$$
$$= \frac{23.19}{2}$$
$$= 11.595$$

$$\text{Upper quartile} = \frac{12.05+12.05}{2}$$
$$= \frac{24.10}{2}$$
$$= 12.05$$

The lowest value is 10.44.

The highest value is 12.63.



a. The group that took ten deep breaths has the fastest runner. The fastest runner took only 10.44 s to run 100 m.

b. The group that did not take the deep breaths has the higher median.

c. The group that took ten deep breaths is faster.

d. If a runner takes ten deep breaths before running, this athlete will run faster than ones that do not take ten deep breaths.

Activity 2

Identify the vocabulary of probability as it applies to the likelihood of the occurrence of an event within a sample.

1. When one letter is selected at random from the word **dimensional**, the specified probabilities are as follows:

   a. $P(m) = \frac{1}{11}$

   b. $P(i) = \frac{2}{11}$

   c. The letters a, e, i, o, and u are vowels. Five of the letters in the word **dimensional** are vowels.

   $P(vowel) = \frac{5}{11}$

d. Consonants are the letters which are not vowels.

$$P(consonant) = \frac{6}{11}$$

e. $P(o) = \frac{1}{11}$

f. Add the individual probabilities here.

$$P(d \text{ or } l) = \frac{2}{11}$$

2. The specific probabilities are as follows:

a. $P(1) = \frac{1}{3000}$

b. $P(10) = \frac{10}{3000}$
$= \frac{1}{300}$

c. $P(300) = \frac{300}{3000}$
$= \frac{1}{10}$

d. $P(0) = \frac{0}{3000}$
$= 0$

e. $P(3000) = \frac{3000}{3000}$
$= 1$

3. Random slips of paper, numbered 1 to 50, are placed in a box. The specific probabilities are as follows:

a. $P(2) = \frac{1}{50}$

b. $P(even\ number) = \frac{25}{50}$
$= \frac{1}{2}$

c. Numbers such as 3, 6, 9, 12, 15, 18, and so on are divisible by 3. There are sixteen numbers that are divisible by 3.

$$P(number\ divisible\ by\ 3) = \frac{16}{50}$$
$$= \frac{8}{25}$$

d. $P(33) = \frac{1}{50}$

4. A wheel is divided into twelve equal sections which are numbered 1 to 12, inclusive. The specific probabilities are as follows:

a. $P(5) = \frac{1}{12}$

b. Numbers such as 1, 3, 5, 7, 9, 11, and so on are odd numbers.

$$P(odd\ number) = \frac{6}{12}$$
$$= \frac{1}{2}$$

c. A prime number has exactly two factors which are 1 and the number itself. The numbers 2, 3, 5, 6, 11, and so on are prime numbers.

$$P(\text{prime number}) = \frac{5}{12}$$

d. $P(\text{greater than } 9) = \frac{3}{12}$

$$= \frac{1}{4}$$

Activity 3

Make a probability statement, and show it relates to a sample and the population.

1.

Blood Type	Percentage of Canucklandians
O	45%
A	40%
B	11%
AB	4%

a. i. There is a 45% probability of selecting someone with Type O blood.

ii. The probability of selecting someone with Type A or Type B blood will be 40% + 11% or 51%.

iii. The probability of selecting someone that does not have Type B blood will be 100% − 11% or 89%.

iv. The probability of selecting someone that does not have Type A nor Type AB blood will be 45% + 11% or 56%.

b. You would expect 45% of 100 000 or 45 000 people to have Type O blood from a sample of 100 000 people.

c. $\frac{4}{40} = \frac{40}{x}$

$4x = 40 \times 40$

$4x = 1600$

$\frac{4x}{4} = \frac{1600}{4}$

$x = 400$

You would expect them to keep 400 bottles of Type A on hand.

2. a. The probability that a man smokes is 17.7%.

b. The probability that a woman is a nonsmoker is 35.4% + 17.3% or 52.7%.

c. A man is less likely to smoke.

d. It appears that men find it easier to quit smoking. However, there may be other factors besides determination that lead to men quitting this habit.

Activity 4

Comment on the confidence of predictions that are based on the analysis of a sample.

1. a. This would be a reasonable cause-and-effect relationship only if the oranges were found to be infected.

   b. This would not be considered a reasonable cause-and-effect relationship. A more suitable cause-and-effect relationship would be the quality of the editor's work and the salary.

   c. This would be considered a good cause-and-effect relationship.

2. a. You might feel fairly confident that as the number of reporters increases, the annual sales of the newspaper would also increase. Readers appreciate full coverage of various events, especially sports.

   b. You should know the average speeds of the other cars in the race before you can make an accurate prediction.

   c. This prediction is most likely correct. However, some people may be overweight in their early teens, but they may lose mass in their later teens while they continue to increase in height.

3. Would you feel that you should insult somebody if your friends were insulting that person?

4. Drinking and driving are sensitive subjects for teenagers and parents. On such subjects, people tend to remember what they want to remember, or report what they know they should have done, whether they did it or not.

5. The evidence does suggest that people who eat potatoes encounter a variety of problems. This does not mean there is a cause-and-effect relation between the two variables. It is extremely unlikely that increases in potato consumption would cause increased problems.

6. The best sample would be 13 out of 100 students (part c). The sample size of 100 students is larger than the other sample sizes. The larger the sample size is, the more accurate the survey will be.

Extra Help

1. a. $P(5) = \frac{1}{6}$

   b. $P(1 \text{ or } 2) = \frac{2}{6}$

   $= \frac{1}{3}$

2. a. $P(7) = \frac{1}{12}$

   b. $P(3,\ 4,\ \text{or } 7) = \frac{3}{12}$

   $= \frac{1}{4}$

3. a. There are four cards with the number seven on them.

$$P(7) = \frac{\mathbf{4}}{54}$$
$$= \frac{2}{27}$$

b. There are 26 red cards in a deck of cards.

$$P(red) = \frac{\mathbf{26}}{54}$$
$$= \frac{13}{27}$$

c. There are four cards with the number 2 on them.
There are four cards with the number 5 on them.
There are four cards with a king on them.

$4 + 4 + 4 = \mathbf{12}$

$$P(2,\ 5,\ or\ king) = \frac{\mathbf{12}}{54}$$
$$= \frac{12 \div 6}{54 \div 6}$$
$$= \frac{2}{9}$$

4. a. $P(green) = \frac{2}{38}$
$= \frac{1}{19}$

b. $P(even) = \frac{18}{38}$
$= \frac{9}{19}$

c. $P(37) = \frac{0}{38}$
$= 0$

Extensions

1. a. $_{49}C_5 = \frac{49!}{5!44!}$
$= \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44!}{5 \times 4 \times 3 \times 2 \times 1 \times 44!}$
$= 1\,906\,884$

Chances of winning are 1 in 1 906 884.

b. $_{49}C_4 = \frac{49!}{4!45!}$
$= \frac{49 \times 48 \times 47 \times 46 \times 45!}{4 \times 3 \times 2 \times 1 \times 45!}$
$= 211\,876$

Chances of winning are 1 in 211 876.

c. $_{49}C_3 = \frac{49!}{3!46!}$
$= \frac{49 \times 48 \times 47 \times 46!}{3 \times 2 \times 1 \times 46!}$
$= 18\,424$

Chances of winning are 1 in 18 424.

d. ${}_{49}C_2 = \frac{49!}{2!47!}$

$= \frac{49 \times 48 \times 47!}{2 \times 1 \times 47!}$

$= 1176$

Chances of winning are 1 in 1176.

e. ${}_{49}C_1 = \frac{49!}{1!48!}$

$= \frac{49 \times 48!}{1 \times 48!}$

$= 49$

Chances of winning are 1 in 49.

f. ${}_{36}C_6 = \frac{36!}{6!30!}$

$= \frac{36 \times 35 \times 34 \times 32 \times 31 \times 30!}{6 \times 5 \times 4 \times 3 \times 2 \times 1 \times 30!}$

$= 1\,947\,792$

Chances of winning are 1 in 1 947 792.

g. ${}_{36}C_5 = \frac{36!}{5!31!}$

$= \frac{36 \times 35 \times 34 \times 32 \times 31!}{5 \times 4 \times 3 \times 2 \times 1 \times 31!}$

$= 376\,992$

Chances of winning are 1 in 376 992.

h. ${}_{36}C_4 = \frac{36!}{4!32!}$

$= \frac{36 \times 35 \times 34 \times 32!}{4 \times 3 \times 2 \times 1 \times 32!}$

$= 58\,905$

Chances of winning are 1 in 58 905.

i. ${}_{36}C_3 = \frac{36!}{3!33!}$

$= \frac{36 \times 35 \times 34 \times 33!}{3 \times 2 \times 1 \times 33!}$

$= 7140$

Chances of winning are 1 in 630.

j. ${}_{36}C_2 = \frac{36!}{2!34!}$

$= \frac{36 \times 35 \times 34!}{2 \times 1 \times 34!}$

$= 630$

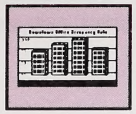
Chances of winning are 1 in 630.

k. ${}_{36}C_1 = \frac{36!}{2!35!}$

$= \frac{36 \times 35!}{1 \times 35!}$

$= 36$

Chances of winning are 1 in 36.



Appendix B
Charts and Tables

Random Number Table

35707	99631	50314	74845	85466	77160	13396	97957	38623	96977
86512	85588	09766	52143	71897	66545	15729	27928	46327	38631
72986	95059	07270	18649	32027	71748	57654	10000	90955	55843
40407	21448	04852	94168	72872	25581	73139	60361	98089	24619
97623	03970	95838	14590	28791	56352	06504	95647	74485	47450
12931	37876	29269	19725	89112	51315	67337	66031	97841	87584
76852	60441	25374	28623	68032	79526	05997	87262	66542	71412
43048	04699	27298	28814	83771	67936	62421	21662	40143	15528
05466	43248	64464	13875	62605	99895	48802	65542	88966	38133
36660	19786	24621	29140	80520	71395	64846	46572	02491	94737
38659	66593	94638	66392	68025	91237	20271	07365	23090	53974
16108	45402	37471	67737	65937	22499	85467	01343	01917	42677
33601	11584	53052	50467	45406	38934	21232	88668	39984	48445
31495	10757	70615	32518	45347	99045	57734	36677	76232	88758
95862	19462	30958	85896	18019	22568	83371	27519	61790	22472
72734	08792	40973	35611	20831	54397	13462	64815	72886	50159
22054	05972	54622	10103	77763	56640	12431	92943	32903	83663
91503	54775	51872	75049	18888	43814	75098	68176	80417	61984
01767	40063	87729	35136	77650	03082	07036	73327	12751	84966
30938	95871	71765	40501	40021	16645	20078	03882	21707	84231
40253	53761	67317	60197	02202	28898	75365	10862	67819	96110
32588	00285	90403	19201	20289	40211	41272	84515	53579	70146
62529	74280	40786	92142	13617	93779	69626	93033	12884	61036
96510	45403	20481	80390	42276	70865	33189	34901	26875	03086
21379	40908	89505	97462	20272	47097	38770	33245	11617	73057

continued on the next page . . .

. . . continued from the previous page

40193	88695	28920	21750	06706	71285	96608	87136	73238	06157
47154	07365	23458	82686	09098	93207	98230	36466	35864	91161
87785	11770	85178	59107	23310	76105	73137	91425	26589	03431
13295	19765	48818	77455	03117	01579	25364	20596	57147	43118
41788	88658	08143	05883	37398	27185	45892	25792	25740	95954
92424	64761	72493	45240	06825	57346	06523	12256	53233	47478
98283	79106	77804	35730	16476	98485	98344	91629	31957	48311
36835	02353	44028	93405	76994	45671	54725	46335	86371	97438
72322	19997	06990	30594	08263	13721	03120	53972	89987	68494
33824	71247	81058	39665	60064	61758	25089	08457	77676	78013
40941	55744	78579	86200	32169	96371	45863	00323	89017	60422
49356	20937	41113	42266	51739	86743	99773	13728	44681	01624
42613	78589	24064	60727	80371	84143	99454	10346	36389	49130
43137	17814	67723	49292	24878	27495	86833	20923	27902	60421
47630	78680	84415	67676	42706	39712	44893	88310	51209	46268
55097	10748	45333	97392	01096	81909	39728	32045	22892	32842
19496	76374	88100	96925	18751	10990	09233	46577	55151	49124
76377	09879	54970	88435	36388	29102	44923	35115	42427	23398
09139	10645	66977	14072	50178	79673	65986	20978	34575	81497
888898	41005	24053	43850	95621	91538	17429	86152	62651	65874
10876	09733	14210	95458	24314	00882	44652	47053	19705	12235
64694	67760	58938	24303	34011	43845	35870	87212	00877	54815
01110	42704	24131	76252	42281	04438	67852	09537	70115	48284
51433	35043	21918	72430	24797	13563	74129	22357	47929	33868
17964	05742	87740	07619	24903	77440	06017	20795	42251	65850

3 3286 50158 4284

L.R.D.C.
Producer

Mathematics 10

Student Module
Unit 8

1992